PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A01N 61/00, 37/18, C07H 21/02, 21/04

A1

(11) International Publication Number: **WO 99/63828**

(43) International Publication Date: 16 December 1999 (16.12.99)

(21) International Application Number: PCT/US99/12910

(22) International Filing Date: 8 June 1999 (08.06.99)

(30) Priority Data:

60/088,521	8 June 1998 (08.06.98)	US
60/134,752	18 May 1999 (18.05.99)	US

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(71) Applicant: EMORY UNIVERSITY [US/US]; 2009 Ridgewood Drive, Atlanta, GA 30322 (US).

(72) Inventors: DOETSCH, Paul, W.; 1978 Castleway Lane, N.E., Atlanta, GA 30322 (US). KAUR, Balveen; 1538 Shoup Court #4, Atlanta, GA 30033 (US). AVERY, Angela, M.; 3078 Clairmont Road #722, Atlanta, GA 30329 (US).

(74) Agents: FERBER, Donna, M. et al.; Greenlee, Winner and Sullivan, P.C., Suite 201, 5370 Manhattan Circle, Boulder, CO 80303 (US).

(54) Title: BROAD SPECIFICITY DNA DAMAGE ENDONUCLEASE

(57) Abstract

The present disclosure describes DNA damage endonucleases which exhibit broad specificity with respect to the types of structural aberrations in double stranded DNA. These enzymes recognize double stranded DNA with distortions in structure, wherein the distortions result from photoproducts, alkylation, intercalation, abasic sites, mismatched base pairs, cisplatin adducts and inappropriately incorporated bases (for example, 8–oxoguanine, inosine, xanthine, among others). The UVDE (Uvelp) of *Schizosaccharomyces pombe*, certain truncated forms of that UVDE (lacking from about 100 to about 250 amino acids of N–terminal sequence) and certain endonucleases from *Homo sapiens, Neurospora crassa, Bacillus subtilis*, and from *Deinococcus radiodurans*. The present disclosure further provides methods for cleaving double stranded DNA having structural distortions as set forth herein using the exemplified endonucleases or their stable, functional truncated derivatives.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

BROAD SPECIFICITY DNA DAMAGE ENDONUCLEASE

ACKNOWLEDGMENT OF FEDERAL RESEARCH SUPPORT

This invention was made, at least in part, with funding from the National Institutes of Health (Grant Nos. CA 55896, AR 42687 and CA 73041), and the National Cancer Institute. Accordingly, the United States Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

The field of the present invention is the area of DNA repair enzymes. In particular, the invention concerns the identification of stable ultraviolet DNA endonuclease polypeptide fragments, their nucleotide sequences and recombinant host cells and methods for producing them and for using them in DNA repair processes.

Cellular exposure to ultraviolet radiation (UV) results in numerous detrimental effects including cell death, mutation and neoplastic transformation. Studies indicate that some of these deleterious effects are due to the formation of two major classes of bipyrimidine DNA photoproducts, cyclobutane pyrimidine dimers (CPDs) and (6-4) photoproducts (6-4 PPs). (Friedberg et al. [1995] in *DNA Repair and Mutagenesis*, pp. 24-31, Am. Soc. Microbiol., Washington, D.C.).

Organisms have evolved several different pathways for removing CPDs and 6-4 PPs from cellular DNA (Friedberg et al. [1995] *supra*; Brash et al. [1991] *Proc. Natl. Acad. Sci. U.S.A.* **88**10124-10128). These pathways include direct reversal and various excision repair pathways which can be highly specific or nonspecific for CPDs and 6-4 PPs. For example, DNA photolyases specific for either CPDs or 6-4 PPs have been found in a variety of species and restore the photoproduct bases back to their original undamaged states (Rubert, C.S. [1975] *Basic Life Sci.* **5A**:73-87; Kim et al. [1994] *J. Biol. Chem.* **269**:8535-8540; Sancar, G.B. [1990] *Mutat. Res.* **236**:147-160). Excision repair has been traditionally divided into either base excision repair (BER) or nucleotide excision repair (NER) pathways, which are mediated by separate sets of proteins but which both are comprised of DNA incision, lesion removal, gap-filling and ligation reactions (Sancar, A. [1994] *Science* **266**:1954-19560; Sancar, A. and Tang, M.S. [1993] *Photochem. Photobiol.* **57**:905-921). BER N-glycosylase/AP lyases specific for CPDs cleave the N-glycosidic bond of the CPD 5' pyrimidine and then cleave the phosphodiester backbone at the abasic site via a β-lyase mechanism, and have been found in several species including T4 phage-infected *Escherichia coli*, *Micrococcus luteus*, and *Saccharomyces cerevisiae* (Nakabeppu, Y. et al. [1982] *J. Biol. Chem.* **257**:2556-2562; Grafstrom, R.H. et al. [1982] *J. Biol. Chem.* **257**:13465-13474; Hamilton, K.K. et al. [1992] *Nature* **356**:725-728). NER is a widely distributed, lesion non-specific repair pathway which orchestrates DNA damage removal via a dual incision reaction upstream and downstream from the damage site, releasing an oligonucleotide containing the damage and subsequent gap filling and ligation reactions (Sancar and Tang [1993] *supra*).

Recently, an alternative excision repair pathway initiated by a direct acting nuclease which recognizes and cleaves DNA containing CPDs or 6-4 PPs immediately 5' to the photoproduct site has been described (Bowman, K.K. et al. [1994] *Nucleic. Acids Res.* **22**:3026-3032; Freyer, G.A. et al. [1995] *Mol. Cell. Biol.* **15**:4572-4577; Doetsch, P.W. [1995] *Trends Biochem. Sci.* **20**:384-386; Davey, S. et al. [1997] *Nucleic Acids Res.* **25**:1002-1008; Yajima, H. et al. [1995] *EMBO J.* **14**:2393-2399; Yonemasu, R. et al. [1997] *Nucleic Acids Res.* **25**:1553-1558; Takao, M. et al. [1996] *Nucleic Acids Res.* **24**:1267-1271). The initiating enzyme has been termed UV damage endonuclease (UVDE, now termed Uve1p). Homologs of UVDE have been found in *Schizosaccharomyces pombe*,

Neurospora crassa and *Bacillus subtilis* (Yajima et al. [1995] *supra*; Yonemasu et al. [1997] *supra*; Takao et al. [1996] *supra*). The Uve1p homologs from these three species have been cloned, sequenced and confer increased UV resistance when introduced into UV-sensitive strains of *E. coli*, *S. cerevisiae*, and human cells (Yajima et al. [1995] *supra*; Takao et al. [1996] *supra*). In *S. pombe* Uve1p is encoded by the *uve1+* gene. However, because of the apparently unstable nature of partially purified full-length and some truncated UVDE derivatives, UVDE enzymes have been relatively poorly characterized and are of limited use (Takao et al. [1996] *supra*).

Because of the increasing and widespread incidence of skin cancers throughout the world and due to the reported inherent instability of various types of partially purified full-length and truncated UVDE derivatives, there is a long felt need for the isolation and purification of stable UVDE products, especially for use in skin care and medicinal formulations.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide purified stable UVDE (Uve1p), polypeptide fragments which retain high levels of activity, particularly those from the *Schizosaccharomyces pombe* enzyme. In a specific embodiment, the polypeptide fragment is Δ228-UVDE, which contains a 288 amino-acid deletion of the N-terminal region of the *S. pombe uve1+* gene product; a second specific embodiment is the fusion protein GST-Δ288-UVDE. The DNA sequence encoding GST-full-length UVDE from *S. pombe* is given in SEQ ID NO:1. The deduced amino acid sequence of full-length UVDE is given in SEQ ID NO:2. The DNA sequence encoding Δ228-UVDE is given in SEQ ID NO:3. The deduced amino acid sequence of Δ228-UVDE is given in SEQ ID NO:4. The DNA coding sequence and deduced amino acid sequence for GST-Δ228-UVDE are given in SEQ ID NO:5 and SEQ ID NO: 6, respectively. Also encompassed within the present invention are truncated UVDE proteins wherein the truncation is from about position 100 to about position 250 with reference to SEQ ID NO:2, and wherein the truncated proteins are stable in substantially pure form.

Also within the scope of the present invention are nucleic acid molecules encoding such polypeptide fragments and recombinant cells, tissues and animals containing such nucleic acids or polypeptide fragments, antibodies to the polypeptide fragments, assays utilizing the polypeptide fragments, pharmaceutical and/or cosmetic preparations containing the polypeptide fragments and methods relating to all of the foregoing.

A specifically exemplified embodiment of the invention is an isolated, enriched, or purified nucleic acid molecule encoding Δ228-UVDE. Another exemplified embodiment is an isolated, enriched or purified nucleic acid molecule encoding GST-Δ228-UVDE.

In a specifically exemplified embodiment, the isolated nucleic acid comprises, consists essentially of, or consists of a nucleic acid sequence set forth in SEQ ID NO:3 or SEQ ID NO:5.

In another embodiment, the invention encompasses a recombinant cell containing a nucleic acid molecule encoding Δ228-UVDE or GST-Δ228-UVDE. The recombinant nucleic acid may contain a sequence set forth in SEQ ID NO:3 or SEQ ID NO:5, a synonymous coding sequence or a functional derivative of SEQ ID NO:3 or SEQ ID NO:5. In such cells, the Δ228-UVDE coding sequence is generally expressed under the control of heterologous regulatory elements including a heterologous promoter that is not normally coupled transcriptionally to the coding sequence for the UVDE polypeptide in its native state.

In yet another aspect, the invention relates to a nucleic acid vector comprising a nucleotide sequence encoding Δ228-UVDE or GST-Δ228-UVDE and transcription and translation control sequences effective to initiate transcription and subsequent protein synthesis in a host cell. Where a GST full length or truncated derivative is expressed, the GST portion is desirably removed (after affinity purification) by protease cleavage, for example using thrombin.

It is yet another aspect of the invention to provide a method for isolating, enriching or purifying the polypeptide termed Δ228-UVDE.

In yet another aspect, the invention features an antibody (e.g., a monoclonal or polyclonal antibody) having specific binding affinity to a UVDE polypeptide fragment. By "specific binding affinity" is meant that the antibody binds to UVDE polypeptides with greater affinity than it binds to other polypeptides under specified conditions.

Antibodies having specific binding affinity to a UVDE polypeptide fragment may be used in methods for detecting the presence and/or amount of a truncated UVDE polypeptide in a sample by contacting the sample with the antibody under conditions such that an immunocomplex forms and detecting the presence and/or amount of the antibody conjugated to the UVDE polypeptide. Kits for performing such methods may be constructed to include a first container having a conjugate of a binding partner of the antibody and a label, for example, a radioisotope or other means of detection as well known to the art.

Another embodiment of the invention features a hybridoma which produces an antibody having specific binding affinity to a UVDE polypeptide fragment. By "hybridoma" is meant an immortalized cell line which is capable of secreting an antibody, for example a Δ228-UVDE specific antibody. In preferred embodiments, the UVDE specific antibody comprises a sequence of amino acids that is able to specifically bind Δ288-UVDE. Alternatively, a GST-tag specific antibody or labeled ligand could be used to determine the presence of or quantitate a GST-Δ228-UVDE polypeptide, especially in formulations *ex vivo*.

The present invention further provides methods for cleaving DNA molecules at positions with structural distortions, wherein the DNA is cleaved in the vicinity of the distortion by a stable truncated UVDE protein of the present invention. The structural distortion can result from mismatch at the site of the distortion in a double-stranded DNA molecule, from UV damage or from other damage to DNA due to chemical reaction, for example, with an alkylating or depurination agent or due to damage due to UV irradiation, ionizing radiation or other irradiation damage. The stable truncated UVDE proteins can be supplied in substantially pure form for *in vitro* reactions or they can be supplied for *in vivo* reactions, including but not limited to compositions for topical application (in the form or of an ointment, salve, cream, lotion, liquid or transdermal patch) in pharmaceutical

compositions for internal use (to be administered by intraperitoneal, intradermal, subcutaneous, intravenous or intramuscular injection). The stable truncated UVDE derivatives of the present invention repair a wide variety of mismatch and DNA damage. The cleavage of a double stranded DNA molecule having structural distortion due to nucleotide mispairing (mismatch) or due to DNA damage by a stable truncated UVDE derivative of the present invention can be used to advantage in a relatively simple assay for structural distortion wherein cleavage of a test molecule (i.e., the double stranded DNA molecule which is being screened for damage, mismatch or other structural distortion) is to be detected.

The present invention further provides a method for cleaning a double stranded DNA molecule in which there is a structural distortion. The structural distortion can be due to aberrations including, but not limited to, base pair mismatch, photoproduct formation, alkylation of a nucleotide such that normal Watson-Crick base pairing is disturbed, intercalation between nucleotides of a compound which could be, for example, an acriflavine, an ethidium halide, among others, or a platinum adduct, for example of a cisplatin moiety. The DNA can also contain an abasic site, an inosine, xanthine, 8-oxoguanine residue, among others. The method of the present invention can be employed using the UVDE (Uve1p) protein from *Schizosaccharomyces pombe*, a truncated derivative of the *S. pombe* UVDE (lacking from about 100 to about 250 N-terminal amino acids), the Δ228-UVDE of *S. pombe*, or the *Neurospora crassa*, *Bacillus subtilis*, *Homo sapiens* or *Deinococcus radiodurans* enzymes as set forth herein (see SEQ ID NOs.:36-39). A specifically exemplified truncated UVDE (Δ228) is given in SEQ ID NO:4. DNA containing the structural distortion is contacted with an enzyme (or active truncated derivative) as described above under conditions allowing endonucleolytic cleavage of one strand of the distorted DNA molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A-1C show purification and activity of GST-Δ228-UVDE and Δ228-UVDE. GST-Δ228-UVDE and Δ228-UVDE from overexpressing *S. cerevisiae* DY150 cells were purified by affinity chromatography on glutathione-Sepharose columns. Fig. 1A shows the

purification of GST-Δ228-UVDE. Proteins were visualized on a silver-stained 12% SDS-polyacrylamide gel. Lanes M: protein molecular weight markers (sizes indicated on the right). Lane 1: 0.5 μg of soluble protein (column load) from crude extract of *S. cerevisiae* overexpressing cells. Lane 2: 0.5 μg of unbound protein from affinity column flow through. Lane 3: 1.0 μg of unbound protein from column wash fractions. Lanes 4-8: equal volume (20 μL) loads of column fractions from affinity column bound proteins eluted with glutathione corresponding to 5, 15, 65, 55, and 35 ng of total protein, respectively. Fig. 1B illustrates SDS-PAGE analysis (silver-stained 12% gel) of proteins following reapplication of GST-Δ228-UVDE onto glutathione-Sepharose and on-column thrombin cleavage to remove the GST tag. Lane M: protein molecular weight markers (sizes indicated on left). Lane 1: 100 ng of GST-Δ228-UVDE (column load). Lane 2: 250 ng thrombin reference marker. Lane 3: 250 ng of Δ228-UVDE eluted from column following thrombin cleavage. Lane 4: 400 ng (total protein) of GST-Δ228-UVDE and GST remaining bound to affinity column following thrombin cleavage and elution with glutathione. Arrows indicate the positions of GST-Δ228-UVDE (A, 68.7 kDa), Δ228-UVDE (B, 41.2 kDa), thrombin (C, 37 kDa), and GST (D, 27.5 kDa). Fig. 1C shows activities of GST-Δ228-UVDE and Δ228-UVDE preparations on CPD-30mer. CPD-30mer was incubated with the following preparations of UVDE: crude extract of overexpressing cell containing vector alone (lane 1), GST-Δ228-UVDE (lane 2), FL-UVDE (lane 3), affinity-purified GST alone (lane 4), affinity-purified GST-Δ228-UVDE (lane 5) and affinity-purified Δ228-UVDE (lane 6). Oligonucleotide cleavage products (14mer) corresponding to UVDE-mediated DNA strand scission of CPD-30mer immediately 5' to the CPD site were analyzed on DNA sequencing gels and subjected to autoradiography and phosphorimager analysis.

Fig. 2 shows the effect of salt concentration on UVDE activity. DNA strand scission assays on end-labeled CPD-30mer were carried out with 150 ng of affinity -purified GST-Δ228-UVDE (open circles) or 40 ng of affinity-purified Δ228-UVDE (closed circles) at pH 7.5 and various concentrations of NaCl under otherwise standard reaction conditions for 20 min (Materials and Methods). Extent of DNA strand scission was determined from phosphorimager analysis of gels. Enzyme activity is expressed as a percentage of CPD-30mer cleaved relative to that observed at 100 mM NaCl (defined as 100%).

Fig. 3 illustrates the effect of pH on UVDE activity. DNA strand scission assays on end-labeled CPD-30mer were carried out with 150 ng of affinity -purified GST-Δ228-UVDE (open circles) or 40 ng of affinity-purified Δ228-UVDE (closed circles) under various pH conditions at otherwise standard reaction conditions for 20 minutes (as described herein). Extent of DNA strand scission was determined from phosphorimager analysis of gels and enzyme activity is expressed as a percentage of CPD-30mer cleaved relative to that observed at pH 6.5 (defined as 100%).

Fig. 4 shows the temperature dependence of UVDE activity. DNA strand scission assays on end-labeled CPD-30mer were carried out with 150 ng of affinity -purified GST-Δ228-UVDE (open circles) or 40 ng of affinity-purified Δ228-UVDE (closed circles) at the indicated temperatures under otherwise standard reaction conditions (See the Examples herein below) for 20 minutes. Extent of DNA strand scission was determined from phosphorimager analysis of gels and enzyme activity is expressed as a percentage of CPD-30mer cleaved relative to that observed at 30°C (defined as 100%).

Figs. 5A-5B illustrates kinetic analysis of CPD-30mer cleavage by purified Δ228-UVDE. Δ228-UVDE (5 nM) was reacted with increasing amounts of 5'-end-labeled CPD-30mer and analyzed for DNA strand scission as described in the Examples. Fig. 5A is a plot of reaction rate (Rate) vs substrate concentration using the mean ± standard deviation from three separate experiments. Curve shown is the best fit to the Michaelis-Menten equation of the averaged data. Fig. 5B is a Lineweaver-Burk plot of the kinetic data.

Fig. 6A-6B show sites of Uve1p cleavage of CPD containing substrates. Various Uve1p preparations were incubated with 5' or 3' end-labeled (*) cs-CPD-30mer. Cleavage products corresponding to Uve1p-mediated strand scission of cs-CPD-30mer were visualized on a DNA sequencing-type gel. Fig. 6A: 5' end labeled cs-CPD-30mer duplex was incubated with buffer only (lane 1), an extract of cells over-expressing GΔ228-Uve1p (5 μg) (lane 2), affinity-purified GΔ228-Uve1p (lane 3) and affinity-purified Δ228-Uve1p (50 ng of each) (lane 4) and affinity-purified GST alone (2 μg) (lane 5). Fig. 6B: 3' end labeled cs-CPD-30mer duplex was incubated with the same Uve1p preparations. Order of lanes is the same as

for Fig. 6A. Arrows **a** and **b** indicate the primary and secondary cleavage sites. The photoproduct (T^T corresponds to CPD) containing section of cs-CPD-30mer is shown at the bottom of the figure. For simplicity the complementary strand is not shown.

Fig. 7A-7D demonstrate that GΔ228-Uve1p recognizes 12 different base mismatch combinations. The 3' end labeled oligo series X/Y-31mer (sequence given at bottom, asterisk indicates labeled strand and labeled terminus) was utilized to assess Uve1p cleavage activity on 16 different base pair and base mispair combinations (Table 1B). Base mispairs indicated above numbered lanes with asterisks denoting base on the labeled strand for G-series (Fig. 7A), A-series (Fig. 7B), C-series (Fig. 7C) and T-series (Fig. 7D) treated with purified GΔ228-Uve1p (odd lanes) or mock reactions (even lanes). Reaction products were analyzed on DNA sequencing-type gels. Arrows indicate Uve1p cleavage sites immediately (arrow a), one (arrow b), and two (arrow c) nucleotides 5' to the mismatch site. G and C + T base-specific chemical cleavage DNA sequencing ladders were run in adjacent lanes as nucleotide position markers.

Figs. 8A-8E show Uve1p activity on bipyrimidine UV induced photoproducts. To determine if Uve1p was capable of recognizing a broad spectrum of UV induced photoproducts, crude extracts from cells expressing GΔ228-Uve1p (lane 1) and G-Uve1p (lane 2) (5 μg of each), and affinity-purified Δ228-Uve1p (lane 3) and GΔ228-Uve1p (lane 4) (50 ng of each) were incubated with the following 5' end-labeled (*) duplex oligonucleotide substrates (Fig. 8A) cs-CPD-49mer, (Fig. 8B) 6-4PP-49mer, (Fig. 8C) tsI-CPD-49mer, (Fig. 8D) tsII-CPD-49mer, and (Fig. 8E) Dewar-49mer. The UV photoproduct (T^T) containing section of the sequence is shown at the bottom of the figure. Arrows **a** and **b** indicate the major and minor products formed by Uve1p mediated cleavage. Arrow **uc** indicates the uncleaved substrate. The sequence of the complementary strand is omitted.

Fig. 9 shows Uve1p activity on a platinum-DNA GG diadduct-containing substrate. Affinity-purified GΔ228-Uve1p (lane 4) and Δ228-Uve1p (1-2 μg) (lane 5) were incubated with 5' end-labeled duplex (*) Pt-Gg-32mer. This substrate was also incubated with buffer alone (lane 2), *E. coli* exonuclease III (150 units (Promega)) (lane 3) and affinity-purified

GST (2 μg) (lane 6). Maxam and Gilbert sequencing (lane 1) of the oligonucleotide was carried out to identify the site of cleavage. Arrows **c** and **d** indicate the major and minor cleavage sites, respectively. The platinum-DNA GG diadduct containing section of the substrate is shown at the bottom of the figure. The sequence of the complementary strand is omitted.

Fig. 10A shows cleavage of an oligonucleotide substrate containing an AP site by Uve1p. To investigate if Uve1p was capable of cleaving an abasic site in a hydrolytic manner, we prepared a 5' end-labeled (*) abasic substrate, AP-37mer, and incubated this substrate with buffer alone (lane 1), *E. coli* endonuclease III (AP lyase, lane 2), affinity-purified GΔ228-Uve1p and Δ288-Uve1p (2 μg of each) (lanes 3 and 4), extracts of cells over-expressing GΔ288-Uve1p (5 μg) (lane 5), *E. coli* endonuclease IV (hydrolytic AP endonuclease, lane 6) and purified recombinant GST (2 μg) (lane 7). Fig. 10B demonstrates competitive inhibition of AP site recognition and cleavage. To demonstrate that the products generated are as a result of Uve1p-mediated cleavage at the AP site, AP-37mer was incubated with buffer alone (lane 1), *E. coli* endonuclease IV (lane 2), and affinity-purified GΔ228-Uve1p (2 μg) (lane 3) with 10X and 40X unlabeled cs-CPD-30mer (lanes 4 and 5, respectively) and 10X and 40X unlabeled UD-37mer (lanes 6 and 7, respectively). Arrows **a** and **b** indicate the primary and secondary Uve1p-mediated cleavage products, respectively. Arrow **uc** indicates the uncleaved substrate. A portion of the sequence of the AP substrate is shown at the bottom of the figure. S corresponds to deoxyribose and p corresponds to phosphate. The location of the cleavage sites of endonuclease III (E_{III}) and endonuclease IV (E_{IV}) are also indicated. For simplicity the complementary strand is omitted from the figure.

Figs. 11A-11B characterize the Uve1p-generated DNA strand scission products and activity of full-length Uve1p. Fig. 11A: Analysis of 5' termini of Uve1p-generated DNA cleavage products with *CX/AY-31mer. 3' end labeled oligo with C/A mismatch (sequence on bottom) reacted with GΔ228-Uve1p and then further treated with PNK or CIP as indicated in the (+) and (-) lanes. Lane 1 is buffer treatment only. Arrows a and b indicate sites of Uve1p cleavage. Fig. 11B: Full length Uve1p possesses mismatch endonuclease activity. 5' end labeled duplex *CX/AY-31mer was incubated with crude extracts of cells expressing

either full-length, GST-tagged Uve1p (GFL-Uve1p) (lane 1), trucated Uve1p (GΔ228-Uve1p) (lane 2), cells expressing the GST tag alone (lane 3) or with *E. coli* endonuclease V, a known mismatch endonuclease (lane 4). Arrows indicate cleavage sites immediately (arrow a) and one nucleotide 5' to the mismatch site. Arrow V indicates *E. coli* endonuclease V cleavage 3' to the mismatch site and was used as a position reference. Bands below arrows (indicated by asterisks) correspond to shortened products due to a weak 5' to 3' exonuclease activity present in the Uve1p preparations.

Fig. 12 shows that GΔ228-Uve1p mismatch endonuclease and GΔ228-Uve1p UV photoproduct endonuclease compete for the same substrates. GΔ228-Uve1p was incubated with 3'-end-labeled duplex *CX/AY-31mer (Table 1) in the presence of increasing amounts of unlabeled duplex CPD-30mer (squares) or duplex GX/CY-31mer (triangles) or duplex CX/AY-31 mer (circles). The Uve1p-mediated DNA cleavage products were analyzed on DNA sequencing gels, and the extent of strand scission was quantified by PhosphorImager analysis. Uve1p activity is expressed as the percentage of the cleavage observed relative to that observed in the absence of any competitor (defined as 100% activity). The error bars indicate the mean ± standard deviation from three separate experiments.

Figs. 13A-13B show that Uve1p incises only one strand of a duplex containing a base mismatch. Fig. 13A shows 3'-end-labeled *CX/AY-41mer incubated with restriction enzyme *Dde*I (lane 1), GΔ228-Uve1p (lane 2), or buffer (lane 3). The reaction products were analyzed on a nondenaturing gel as described below for the presence of DNA double-strand break products (arrow dsb). Arrows b and c indicate the primary cleavage site for Uve1p on this substrate. Fig. 13B shows 3'-end-labeled *CX/AY-41mer or CX/*AY-41mer incubated with GΔ228-Uve1p (+ lanes) or buffer (- lanes) and analyzed on denaturing DNA sequencing-type gels. Arrows b and c indicate positions of major Uve1p cleavage events relative to the mismatched base (asterisk) position. G + A and C + T base-specific sequencing ladders are included in outside lanes as nucleotide position markers.

DETAILED DESCRIPTION OF THE INVENTION

Abbreviations used in the present specification include the following: aa, amino acid(s); bp, base pair(s); BER, base excision repair; cDNA, DNA complimentary to RNA; CPD, cyclobutane pyrimidine dimer; FL, full-length; GST, glutathione-S-transferase; NER, nucleotide excision repair; PAGE, polyacrylamide gel electrophoresis; PMSF, phenylmethanesulfonyl fluoride, 6-4 PP, (6-4) photoproduct; UVDE or Uve1p, used interchangeably, ultraviolet damage endonuclease; Δ228-UVDE, UVDE truncation product lacking 228 N-terminal amino acids.

By "isolated" in reference to a nucleic acid molecule it is meant a polymer of 14, 17, 21 or more nucleotides covalently linked to each other, including DNA or RNA that is isolated from a natural source or that is chemically synthesized. The isolated nucleic acid molecule of the present invention does not occur in nature. Use of the term "isolated" indicates that a naturally occurring or other nucleic acid molecule has been removed from its normal cellular environment. By the term "purified" in reference to a nucleic acid molecule, absolute purity is not required. Rather, purified indicates that the nucleic acid is more pure than in the natural environment.

A "nucleic acid vector" refers to a single or double stranded circular nucleic acid molecule that can be transfected or transformed into cells and replicate independently or within the host cell genome. A circular double stranded nucleic acid molecule can be linearized by treatment with the appropriate restriction enzymes based on the nucleotide sequences contained in the cloning vector. A nucleic acid molecule of the invention can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together. The nucleic acid molecule can be RNA or DNA.

Many techniques are available to those skilled in the art to facilitate transformation or transfection of the recombinant construct into a prokaryotic or eukaryotic organism. The terms "transformation" and "transfection" refer to methods of inserting an expression construct into a cellular organism. These methods involve a variety of techniques, such as

treating the cells with high concentrations of salt, an electric field, or detergent, to render the host cell competent for uptake of the nucleic acid molecules of interest or liposome-mediated transfection can be employed.

The term "promoter element" describes a nucleotide sequence that is incorporated into a vector which enables transcription, in appropriate cells, of portions of the vector DNA into mRNA. The promoter element precedes the 5' end of the Δ228-UVDE or GST-Δ228-UVDE nucleic acid molecule such that the Δ228-UVDE OR GST-Δ228-UVDE sequence is transcribed into mRNA. Transcription enhancing sequences may also be incorporated in the region upstream of the promoter. mRNA molecules are translated to produce the desired protein(s) within the recombinant cells.

Those skilled in the art would recognize that a nucleic acid vector can contain many other nucleic acid elements besides the promoter element and the Δ228-UVDE or GST-Δ228-UVDE nucleic acid molecule. These other nucleic acid elements include, but are not limited to, origins of replication, ribosomal binding sites, transcription and translation stop signals, nucleic acid sequences encoding drug resistance enzymes or amino acid metabolic enzymes, and nucleic acid sequences encoding secretion signals, periplasm or other localization signals, or signals useful for polypeptide purification.

As used herein, "Δ228-UVDE polypeptide" has an amino acid sequence as given in or substantially similar to the sequence shown in SEQ ID NO:4. A sequence that is substantially similar will preferably have at least 85% identity and most preferably 99-100% identity to the sequence shown in SEQ ID NO:4. Those skilled in the art understand that several readily available computer programs can be used to determine sequence identity with gaps introduced to optimize alignment of sequences being treated as mis-matched amino acids and where the sequence in SEQ ID NO:4 is used as the reference sequence.

As used herein, "GST-Δ228-UVDE polypeptide has an amino acid sequence as given in or substantially similar to the sequence shown in SEQ ID NO:6. A sequence that is substantially similar will preferably have at least 85% identity and most preferably 99-100%

identity to the sequence shown in SEQ ID NO:6. Those skilled in the art understand that several readily available computer programs can be used to determine sequence identity with gaps introduced to optimize alignment of sequences being treated as mis-matched amino acids and where the sequence in SEQ ID NO:6 is used as the reference sequence.

By "isolated" in reference to a polypeptide is meant a polymer of 6, 12, 18 or more amino acids conjugated to each other, including polypeptides that are isolated from a natural source or that are chemically synthesized. The isolated polypeptides of the present invention are unique in the sense that they are not found in a pure or separated state in nature. Use of the term "isolated" indicates that a naturally occurring sequence has been removed from its normal cellular environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. The term does not imply that the sequence is the only amino acid chain present, but that it is essentially free (at least about 90-95% pure) of material naturally associated with it.

The term "purified" in reference to a polypeptide does not require absolute purity (such as a homogeneous preparation); instead, it represents an indication that the polypeptide is relatively purer than in the natural environment. Purification of at least two orders of magnitude, preferably three orders of magnitude, and more preferably four or five orders of magnitude is expressly contemplated, with respect to proteins and other cellular components present in a truncated UVDE-containing composition. The substance is preferably free of contamination at a functionally significant level, for example 90%, 95%, or 99% pure. Based on increases in calculated specific activity, GST-Δ228-UVDE and Δ228-UVDE have been purified 230-fold and 310-fold, respectively. However, based on silver-stained SDS polyacrylamide gel results, it appears that both proteins have been purified nearly to homogeneity (see Fig.1).

As used herein, a "UVDE polypeptide fragment" or "truncated UVDE" has an amino acid sequence that is less than the full-length amino acid sequence shown in SEQ ID NO:2. Also as used herein, UVDE and Uve1p are used synonynously.

In the present context, a "UVDE mutant polypeptide" is a UVDE polypeptide or truncated UVDE which differs from the native or truncated native sequence in that one or more amino acids have been changed, added or deleted. Changes in amino acids may be conservative or non-conservative. By conservative it is meant the substitution of an amino acid for one with similar properties such as charge, hydrophobicity and structure. A UVDE mutant polypeptide of the present invention retains its useful function, i.e., for example, ability to remove cyclobutane pyrimidine dimers and/or (6-4) photoproducts from DNA, and its enzymatic activity is stable in its substantially purified form. The full-length UVDE protein and the truncated derivatives of the present invention recognize a wide variety of DNA damage and distortions to double stranded DNA, as described hereinbelow. The UVDE and truncated UVDE proteins are useful in cleaving double-stranded DNA molecules in which damage including but not limited to abasic sites, photoproducts, cis-platin adducts and a variety of other aberrations also including mismatched base pairing and sites adjacent to and at locations of intercalations (for example with acridine dyes or ethidium bromide, among others, and these proteins, particularly the stable truncated derivatives of the present invention are useful in vivo and/or in vitro for repairing DNA distortions as described herein.

The isolation of genes encoding UVDEs from different organisms has been described previously (Yajima et al. [1995] *supra*; Takao et al. [1996] *supra*). These genes have been cloned by introducing a foreign cDNA library into a repair-deficient *E. coli* strain and selecting for complemented cells by UV irradiation of the transformants. (Yajima et al. [1995] *supra*; Takao et al. [1996] *supra*). Researchers have not characterized full-length UVDEs because they become unstable and lose their activity when purified (Takao et al. [1996] *supra*). This instability makes their use as therapeutic agents problematical.

Because UVDEs can be used for a variety of applications including the treatment and prevention of diseases caused by DNA damage, the inventors sought to discover stable UVDEs. The present inventors have noted that the activity of the full-length UVDE appears relatively stable to storage and freeze-thawing when it is present in crude extracts of either its native *Schizosaccharomyces pombe* or recombinant *Escherichia coli* (see also Takao et al. [1996] *supra*). The present inventors and others have not had success in obtaining

enzymatically active purified UVDE in good yield. The present invention describes the isolation and purification of a polypeptide fragment from *S. pombe* which exhibits superior stability and enzymatic activity than purified full-length UVDE.

The full-length *uvde* gene from *S. pombe* was amplified from a cDNA library by the polymerase chain reaction (PCR) using methods known to those skilled in the art and as described herein. Δ228-UVDE, which contains a deletion of the of the first 228 N-terminal amino acids of full-length UVDE, was prepared using PCR as described herein.

The amplified UVDE gene coding fragments were cloned into the yeast expression vector pYEX4T-1. In pYEX 4T-1, the UVDE-derived polypeptides are expressed in frame with a glutathione-S-transferase (GST) leader sequence to generate a fusion protein of GST linked to the N-terminus of UVDE. The DNA sequence of the GST leader is shown in SEQ ID NO:7. The deduced amino acid sequence of the GST leader is shown in SEQ ID NO:8. Appropriate plasmids containing the DNA fragments in the proper orientation were transformed into *S. cerevisiae*, DY150 cells using the alkali cation method (Ito, H. et al. [1993] *J. Bacteriol.* **153**:163-163). Positive clones were selected and used for protein purification.

Both full-length UVDE and Δ228-UVDE were isolated and purified using glutathione-Sepharose affinity chromatography. Extracts from cells expressing GST-Δ228-UVDE were passed through glutathione-Sepharose columns. GST-Δ228-UVDE which bound to the column was eluted using glutathione. Additionally, Δ228-UVDE was generated by removal of the GST-leader from GST-Δ228-UVDE by treating GST-Δ228-UVDE, which had bound to the glutathione-Sepharose column, with thrombin. Pooled fractions from the affinity purification yielded approximately 1.5 mg of near-homogeneous or homogeneous GST-Δ228-UVDE protein per 500 mL of *S. cerevisiae* cells.

GST-Δ228-UVDE and Δ228-UVDE have electrophoretic mobilities corresponding to protein sizes, as determined by SDS-PAGE, of 68.7 kDa and 41.2 kDa, respectively (Fig. 1A, lanes 4-8; Fig. 1B, lane 3). Both crude and purified preparations of Δ228-UVDE and GST-

Δ228-UVDE retained enzymatic activity on an oligodeoxynucleotide substrate (CPD-30mer) containing a single cis-syn cyclobutane pyrimidine dimer embedded near the center of the sequence (Fig. 1C). In contrast, purified full-length UVDE resulted in a preparation that was not stable in that enzymatic activity was rapidly lost (Fig. 1C, lane 3). Furthermore, purified GST-Δ228-UVDE and Δ228-UVDE are stable when stored at -80°C in 10% glycerol for a period of at least six months with no substantial loss of activity. Preparations of GST-Δ228-UVDE and Δ228-UVDE are resistant to several rounds of freeze-thawing. Surprisingly, both purified GST-Δ228-UVDE and Δ228-UVDE are more stable and have higher enzymatic activity than purified full-length UVDE.

Both truncated forms of UVDE (GST-Δ228-UVDE and Δ228-UVDE) retained high levels of activity over a broad NaCl concentration range (50-300mM) with an optimum around 100mM (Fig. 2). Optimal cleavage of an oligodeoxynucleotide substrate (CPD-30mer) occurred in the presence of 10mM $MgCl_2$ and 1 mM $MnCl_2$. Both GST-Δ228-UVDE and Δ228-UVDE showed optimal cleavage of CPD-30mer at pH 6.0-6.5 with activity sharply declining on either side of this range indicating that the GST tag does not affect the folding and activity of the protein (Fig. 3). The calculated pI values for GST-Δ228-UVDE and Δ228-UVDE are 6.8 and 7.5, respectively.

Under optimal pH, salt and divalent cation conditions, GST-Δ228-UVDE and Δ228-UVDE were found to exhibit a temperature optimum at 30°C (Fig. 4). At 37°C GST-Δ228-UVDE and Δ228-UVDE activities decreased to approximately 85% and 60%, respectively and at 65°C, both truncated versions of UVDE showed a significant decrease in activity.

The kinetic parameters for homogeneous GST-Δ228-UVDE and Δ228-UVDE were determined using the CPD-30mer substrate. Fig. 5 shows that Michaelis-Menten kinetics apply to the CPD-30mer cleavage reactions with Δ228-UVDE. Fig. 5B is a Lineweaver-Burk plot of the kinetic data in Fig. 5A. The apparent K_m for CPD-30mer was calculated to be 49.1 nM ± 7.9 nM for GST-Δ228-UVDE and 74.9 nM ± 3.6 nM for Δ228-UVDE. The V_{max} values (nM min^{-1}) were found to be 2.4 ± 0.13 and 3.9± 0.12 for GST-Δ228-UVDE and

Δ228-UVDE, respectively. The turnover numbers (K_{cat}) were 0.21 ± 0.01 min^{-1} for GST-Δ228-UVDE and 0.9 ± 0.03 min^{-1} for Δ228-UVDE.

Uve1p has been shown to be capable of recognizing both *cis-syn* CPDs (cs-CPD) and 6-4PPs (Bowman et al. [1994] *Nucl. Acids Res.* **22**:3036-3032; Yajima et al. [1995] *EMBO J.* **14**:2393-2399). It is unique in this respect as no other single polypeptide endonuclease is known to recognize both of these UV photoproducts. CPDs and 6-4PPs are the most frequently occurring forms of UV-induced damage, but there are significant differences in the structural distortions induced in DNA by these two lesions. Incorporation of a cs-CPD into duplex DNA causes no significant bending or unwinding of the DNA helix (Rao et al. [1984] *Nucl. Acids Res.* **11**:4789-4807; Wang et al. [1991] *Proc. Natl. Acad. Sci. USA* **88**:9072-9076; Miaskiewicz et al. [1996) J. Am. Chem. Soc. **118**:9156-9163; Jing et al. [1998] *supra*; McAteer et al. [1998] *J. Mol. Biol.* **282**:1013-1032; Kim et al. [1005] *supra*) and destabilizes the duplex by ~1.5 kcal/mol (Jing et al. [1998] *Nucl. Acids Res.* **26**:3845-3853). It has been demonstrated that this relatively small structural distortion allows CPD bases to retain most of their ability to form Watson-Crick hydrogen bonds (Jing et al. [1998] *supra*; Kim et al. [1995] *Photochem. Photobiol.* **62**:44-50). On the other hand, NMR studies have suggested that 6-4PPs bend the DNA to a greater extent than cs-CPDs, and there is a destablization of ~6 kcal/mol in the DNA duplex with a resulting loss of hydrogen bond formation at the 3'-side of the 6-4PP DNA adduct (Kim et al. [1995] *Eur. J. Biochem.* **228**:849-854). The ability of Uve1p to recognize such different structural distortions suggests that it might also recognize other types of DNA damage.

CPDs can occur in DNA in four different isoforms (*cis-syn* I [cs I], *cis-syn* II [cs II], *trans-syn* I [ts I] and *trans-syn* II [ts II]) (Khattak, M.N. and Wang, S.Y. [1972] *Tetrahedron* **28**:945-957). Pyrimidine dimers exist predominately in the cs I form in duplex DNA whereas *trans-syn* (ts) dimers are found primarily in single stranded regions of DNA. 6-4PPs are alkali labile lesions at positions of cytosine (and much less frequently thymine) located 3' to pyrimidine nucleosides (Lippke et al. [1981] *Proc. Natl. Acad. Sci. USA* **78**:3388-3392). 6-4PPs are not stable in sunlight and are converted to their Dewar valence isomers upon exposure to 313 nm light. We have investigated the specificity of Δ228-Uve1p for a series of

UV photoproducts: cs-CPD, ts I-CPD, ts II-CPD, 6-4PP and the Dewar isomer. We also investigated the possibility that Uve1p may recognize other types of non-UV photoproduct DNA damage. We describe the activity of Uve1p on DNA oligonucleotide substrates containing a variety of lesions including a platinum-DNA GG diadduct (Pt-GG), uracil (U), dihydrouracil (DHU), 8-oxoguanine (8-oxoG), abasic sites (AP site), inosine (I), and xanthine (Xn). This collection of substrates contains base lesions that induce a broad range of different DNA structural distortions.

Uve1p isolated from *S. pombe* was first described as catalyzing a single ATP-independent incision event immediately 5' to the UV photoproduct, and generating termini containing 3' hydroxyl and 5' phosphoryl groups (Bowman et al. [1994] *Nucl. Acids Res.* **22**:3026-3032). The purified GΔ288-Uve1p, Δ288-Uve1p and crude cell lysates of recombinant G-Uve1p and GΔ288-Uve1p make an incision directly 5' to CPDs similar to that observed with the native protein. In this study, we have used both 5' and 3' end-labeled duplex CPD-30mer (cs-CPD-30mer) to demonstrate the ability of Uve1p to cleave a CPD-containing substrate at two sites (Fig. 6A-6B). The primary product (arrow a) accounted for approximately 90% of the total product formed and resulted from cleavage immediately 5' to the damage. The second incision site was located one nucleotide upstream and yielded a cleavage product (arrow b), which represented the remaining 10% of the product formed. This minor product is one nucleotide shorter or longer than the primary product depending on whether 5' or 3' end-labeled substrate is being examined. The same cleavage pattern was observed for each different Uve1p preparation used: i.e., crude extracts of cells expressing GΔ228-Uve1p, affinity-purified GΔ228-Uve1p and Δ228-Uve1p (Fig. 2A and 2B, lanes 2, 3 and 4 respectively), as well as extracts of cells expressing GST-Uve1p. No cleavage products were observed when the cs-CPD-30mer substrates were incubated with buffer only, or purified recombinant GST prepared and affinity-purified in an identical manner to the purified Uve1p proteins (Fig. 6A, 6B, lanes 1 and 5 respectively). This control eliminates the possibility that these DNA strand scission products are formed as a result of the presence of trace amounts of non-specific endonuclease contamination. Uve1p recognizes a duplex cs-CPD-containing oligonucleotide substrate and cleaves this substrate at two sites. The primary

site, responsible for 90% of the product, is immediately 5' to the damage and the secondary site (accounting for the remaining 10% of product), is one nucleotide 5' to the site of damage.

Uvelp cleaves both CPDs and 6-4PPs when they are incorporated into oligonucleotide substrates (Bowman et al. [1994] *supra*; Yajima et al. [1995] *EMBO J.* **14**:2393-2399). These lesions induce substantially different distortions in duplex DNA. The ability of native Uvelp to recognize both of these damages prompted us to investigate whether this endonuclease recognized other forms of UV-induced photodamage, as well. In order to determine the substrate range of recombinant Δ228-Uvelp for UV-induced bipyrimidine photoproducts, various Uvelp preparations were incubated with synthetic 49-mer oligonucleotides containing different forms of UV damage (Table 1A). The substrates used in these experiments were 5' end labeled duplex cs-CPD-49mer, tsI-CPD-49mer, tsII-CPD-49mer, 6-4PP-49mer and Dewar-49mer (Fig. 8A). Generally, purified GΔ228-Uvelp and Δ228-Uvelp cleaved all of the bipyrimidine photoproduct substrates in a similar manner with respect to both the site and extent of cleavage. The cleavage pattern observed when crude cell lysates of G-Uvelp and GΔ228-Uvelp were incubated with the substrates was less consistent. Very low levels of product were observed when these extracts were incubated with the Dewar isomer. No cleavage products were detected when the damaged substrates were incubated with buffer alone or purified recombinant GST, demonstrating that no other DNA repair proteins were responsible for the cleavage of the substrate. In addition, incubation of Uvelp with end-labeled undamaged substrate (UD-30mer) did not result in the formation of any cleavage products. We concluded that Uvelp recognizes and cleaves these five UV-induced bipyrimidine photoproducts in a similar manner and that they are substrates for this enzyme. This is the first time that a single protein endonuclease capable of recognizing such a surprisingly broad range of UV-induced photoproducts has been described.

To explore activity on DNA with non-UV-photoproduct diadducts we investigated whether Uvelp recognized an oligonucleotide containing a platinum-DNA lesion. *cis*-Diamminedichloroplatinum(II) (cisplatin) is a widely used antitumor drug that induces several types of mono- and diadducts in DNA. One of the major, biologically relevant

adducts formed results from the coordination of N-7 of two adjacent guanines to platinum to form the intrastrand crosslink *cis*-[$Pt(NH_3)_2${d(GpG)-N7(1),-N7(2)}] (*cis*-PT-GG) (Fig. 9). A 5' end-labeled duplex 32-mer oligonucleotide with a single platinum intrastrand crosslink between positions 16 and 17 (Pt-GG-32mer) (Table 1A) was incubated with either GΔ228-Uve1p or Δ228-Uve1p, and the reaction products were visualized on a DNA sequencing-type gel (Fig. 9). The 3' to 5' exonuclease activity of *E. coli* exonuclease III was used to identify the specific site of cleavage of Uve1p, as a platinum-DNA diadduct will terminate or stall the digestion of the duplex DNA at this site (Royer-Pokora et al. [1981] *Nucl. Acids Res.* **9**:4595-4609; Tullius, T.D. and Lippard, S.J. [1981] *J. Am. Chem. Soc.* **103**:4620-4622). Incubation of 5' end-labeled Pt-GG-32mer with exonuclease III (Fig. 9, lane 3) generates 5' end-labeled oligonucleotide fragments with 3' hydroxyl termini. Maxam and Gilbert sequencing (Fig. 9, lane 1) of the same substrate generates 5' end labeled fragments with 3' phosphoryl termini which consequently migrate faster than the exonuclease III product on DNA sequencing-type gels. (Due to overreaction with hydrazine all of the nucleotides are highlighted in the sequencing lane.) GΔ228-Uve1p cleaved Pt-GG-32mer 5' to the GpG adduct position at two adjacent sites (Fig. 9, lane 4, arrows c and d). The products c) and d) migrate with the exonuclease III products, confirming that they have 3' hydroxyl termini. Comparison with the Maxam and Gilbert sequencing ladder (Fig. 9, lane 1) indicates that the GΔ228-Uve1p-mediated cleavage products are generated by cleavage at sites located two and three nucleotides 5' to the platinum DNA-GG diadduct. The GΔ228-Uve1p-mediated cleavage products were quantified by phosphorimager analysis, and it was determined that cleavage at the primary site c (arrow c) accounted for approximately 90% of the total product formed while cleavage at the secondary site (arrow d) accounted for the remaining 10%. In contrast, Δ228-Uve1p appeared to cleave Pt-GG-32mer only at the primary site c (i.e., two nucleotides 5' to the damage) (Fig. 9, lane 5). When the quantity of protein used and the total amount of product formed is taken into account, the cleavage of Pt-GG-32mer by Uve1p appears at least 100-fold less efficient than the cleavage of the UV-induced photoproducts. Despite this significant decrease in efficiency, Pt-GG-32mer is a substrate for Uve1p, albeit a poor one, and more importantly, Uve1p is capable of recognizing and cleaving a non-UV photoproduct dimer lesion.

Uve1p is active on substrates containing non-bulky DNA damages. The ability of Uve1p to recognize and cleave non-UV photoproduct DNA diadducts prompted us to investigate whether other types of base damage could also be recognized by this versatile endonuclease. These damages included abasic sites (AP sites), uracil (U), dihydrouracil (DHU), inosine (I), xanthine (Xn) and 8-oxoguanine (8-oxoG) (Scheme 1C). For these studies, we utilized 37-mer oligonucleotide substrates with the damages placed near the center of the molecule and within the same DNA sequence context (Table 1B). These oligonucleotides, Ap-37mer, U-37mer, DHU-37mer and 8-oxoG-37mer were incubated with various Uve1p preparations, and the reaction products were analyzed on DNA sequencing-type gels. In addition, 31mer oligonucleotides containing inosine (I-31mer) and xanthine (Xn-31mer) were also tested as potential Uve1p substrates (Table 1A).

Abasic sites (AP sites) arise in DNA from the spontaneous hydrolysis of N-glycosyl bonds and as intermediates in DNA glycosylase-mediated repair of damaged bases (Sakumi, K. and Sekiguchi, M. [1990] *Mutat. Res.* **236**:161-172). AP endonucleases cleave hydrolytically 5' to the site to yield a 3'hydroxyl termini, AP lyases cleave by a β-elimination mechanism leaving a 3'-αβ-unsaturated aldehyde (Spiering, A.L. and Deutsch, W.A. [1981] *J. Biol. Chem.* 261:3222-3228). To determine if Uve1p cleaves AP sites, we incubated affinity-purified GΔ228-Uve1p and Δ228-Uve1p and crude extracts of cells expressing GΔ228-Uve1p with a 5' end-labeled oligonucleotide substrate containing an AP site placed opposite a G residue (AP/G-37mer). The products were analyzed on a DNA sequencing-type gel as before (Fig. 10A, lanes 3, 4 and 5 respectively). *E. coli* endonuclease III (which has an associated AP lyase activity) and *E. coli* endonuclease IV (a hydrolytic AP endonuclease) were used to determine if the cleavage products formed during incubation with Uve1p preparations were due to a β-elimination mechanism or hydrolytic cleavage (Fig. 10A, lanes 2 and 6 respectively). Uve1p recognized the AP site in this oligonucleotide substrate and cleaved it in a similar manner to *E. coli* endonuclease IV. Incubating the Uve1p proteins with an oligonucleotide substrate where the AP site was placed opposite an adenine residue (AP/A-37mer) resulted in no significant change in the amount of cleavage product formed. To further test Uve1p recognition of AP sites, we used unlabeled cs-CPD-30mer as a specific competitor for Uve1p. Addition of 40X unlabeled CPD-30mer to reactions of a 5' end-

labeled AP/G-37mer with the purified GΔ228-Uve1p resulted in an ~60% decrease in the amount of product formed. The addition of 40X unlabeled undamaged 30mer (UD-30mer) had no effect on the amount of product observed. Uve1p is capable of recognizing AP sites, and changing the complementary base has little or no effect on the extent of cleavage.

Uracil lesions can occur in DNA by the spontaneous deamination of a cytosine residue. Dihydrouracil is a pyrimidine photoproduct that is formed by the deamination of cytosine with subsequent ring saturation upon exposure to ionizing radiation under anoxic conditions (Dizdaroglu et al. [1993] *Biochemistry* **45**:12105-12111). To determine if Uve1p recognized uracil and dihydrouracil lesions, we incubated various preparations of Uve1p with 3' end-labeled 37mer oligonucleotides containing uracil and DHU residues placed opposite a G (U/G-37mer, DHU/G-37mer). The results of this set of experiments are summarized in Table 2. Purified GΔ228-Uve1p cleaved U/G-37mer and DHU/G-37mer in a typical Uve1p mediated fashion: immediately 5' to the position of the lesion to form a major product, and again one nucleotide 5' to the damaged site to form a minor product, 90% and 10% of the total Uve1p-mediated cleavage products, respectively.

Persistence of uracil and DHU lesions through replication may lead to the incorporation of adenine residues opposite the damaged base. To examine if Uve1p were equally efficient at recognizing uracil and DHU when they were base paired with an adenine residue, we constructed the substrates U/A-37mer and DHU/A-37mer. The results obtained from the analysis of Uve1p cleavage of these substrates are summarized in Table 2. No Uve1p mediated cleavage products were observed when crude extracts from cells expressing GΔ228-Uve1p and purified GΔ228-Uve1p were incubated with the U/A-37mer. Incubating purified GΔ228-Uve1p with DHU/A-37mer rather than DHU/G-37mer resulted in a 4-fold decrease in the amount of Uve1p-mediated cleavage products observed. To determine whether Uve1p cleaves the complementary strand of these substrates (i.e., U/A-37mer, DHU/A-37mer or U/G-37mer, DHU/G-37mer), we conducted similar experiments with these substrates except that the complementary strand was 3' end-labeled. No cleavage products were observed when these substrates were incubated with purified Uve1p protein preparations. Uve1p recognizes and cleaves uracil and DHU when they are placed opposite a

G (U/G or DHU/G). However, when the lesions are placed in a situation where Watson-Crick hydrogen-bonding is maintained (U/A or DHU/A), Uve1p either fails to recognize the lesion completely (U/A) or the extent of cleavage is significantly decreased (DHU/G).

Uve1p recognizes and cleaves oligonucleotide substrates containing AP sites, uracil and DHU lesions. AP sites appear to be better substrates for Uve1p than uracil or DHU containing oligonucleotides; Uve1p cleaved AP sites at least 10 times more efficiently than uracil containing substrates and twice as efficiently as DHU containing substrates. However, they are all poorer substrates than UV-induced photoproducts. See Table 3 for a summary of the relative efficiency for cleavage by Uve1p on various substrates.

Additionally, the Uve1p preparations were incubated with the following substrates to determine if these lesions were capable of being cleaved by Uve1p: inosine and xanthine placed opposite a T or C (I/T-31mer, I/C-31mer and Xn/T-31mer, Xn/C-31mer), and 8-oxoguanine placed opposite all four bases (8-oxoG/G-37mer, 8-oxoG/A-37mer, 8-oxoG/T-37mer, 8-oxoG/C-37mer). No cleavage of either strand in these duplex substrates was observed.

As discussed hereinabove, because of substantial structural differences between CPDs and 6-4PPs, it was not obvious what features of damaged DNA Uve1p recognizes. One possibility is that Watson-Crick base pairing is disrupted for the 3' pyrimidines in both CPDs and 6-4PPs (Jing et al. [1998] *Nucl. Acids. Res.* **26**:3845-3853), suggesting that Uve1p might target its activity to mispaired bases in duplex DNA. We therefore investigated the ability of purified GΔ228-Uve1p to cleave duplex oligonucleotides containing all possible combinations of single base mispairs embedded within the same flanking sequence context. For these studies, we utilized a collection of mismatch-containing oligonucleotides (series XY-31mer) which were designed so as to generate all possible mismatch combinations (Table 1B). Strands GX, AX, TX and CX were 3' end-labeled and then annealed to strands GY, AY, TY or CY prior to incubation with purified GΔ228-Uve1p. Reaction products were analyzed on DNA sequencing-type gels (See Examples). The ability of GΔ228-Uve1p to cleave all twelve possible mispair combinations is shown in Fig. 7A-7D. No DNA strand

cleavage was observed for duplex substrates containing normal Watson-Crick G/C or A/T base pairs.

The sites of GΔ228-Uve1p-mediated mismatch-specific DNA cleavage were identified in each case by comparing the electrophoretic mobilities of the DNA strand scission products to those of a DNA sequencing ladder obtained by base-specific chemical cleavage. Arrows a, b, and c indicate the DNA strand scission products corresponding to cleavage by GΔ228-Uve1p immediately (position 0), one (position -1) or two (position -2) nucleotides 5' to the site of the mismatch, respectively (Fig. 7A-D). These sites of GΔ228-Uve1p-mediated endonucleolytic cleavage were confirmed in similar experiments employing 5' end-labeled GX, AX, TX and CX strands in the mismatch substrates. In addition, the non-truncated, full-length GFL-Uve1p (in crude cell extracts) recognized and cleaved *CX-AY-31mer in a manner identical to GΔ228-Uve1p (Fig. 11B). The preferred sites of cleavage and the efficiency with which each mismatch is recognized by GΔ228-Uve1p is variable and depends on the type of base mispair that is presented to the enzyme. Within the sequence context examined, GΔ228-Uve1p exhibited strong cleavage at *C/C (asterisk - labeled strand base), *C/A and *G/G sites, moderate cleavage at *G/A, *A/G and *T/G sites, and weak cleavage at *G/T, *A/A, *A/C, *C/T, *T/T and *T/C sites. These differences in the extent of cleavage were reproducible and observed in three separate experiments. These results indicate that the GΔ228-Uve1p mismatch endonuclease activity has a preference for certain base mismatch combinations (e.g. *C/A) over others (e.g. *T/C). However, these experiments do not rule out an effect on cleavage by the sequence(s) flanking the mismatch.

Uve1p has been shown to incise DNA containing CPDs and 6-4PPs directly 5' to the photoproduct site generating products containing 3'-hydroxyl and 5'-phosphoryl groups (Bowman et al. [1994] *supra*). We examined whether similar 3' and 5' termini were produced following Uve1p-mediated cleavage of base mismatch-containing substrates. DNA strand scission products generated by GΔ228-Uve1p cleavage of 3' end-labeled oligo *CX/AY-31mer (CX strand labeled, Table 1B) were further treated with calf intestinal phosphatase (CIP) which removes 5' terminal phosphoryl groups from substrate DNA. The major sites of Uve1p-mediated DNA cleavage relative to the base mispair site were found to be at positions

0 and -1 (Fig. 11A, lane 2). CIP treatment of these DNA cleavage products resulted in species that had retarded electrophoretic mobilities compared to non-CIP-treated DNA cleavage products, indicating a decrease in charge corresponding to removal of 5' terminal phosphoryl groups (Fig. 11A, lanes 2 and 3). In addition, GΔ228-Uve1p mismatch endonuclease-generated DNA cleavage products were resistant to phosphorylation by polynucleotide kinase, an expected result if the 5' termini already contain phosphoryl groups (Fig. 11A, lane 4). Electrophoretic mioibility shift analysis utilizing 5' end-labeled *CX/AY-31mer, terminal deoxyribonucleotidyl transferase (TdT), and α^{32}P-dideoxyATP (ddATP) resulted in addition of a single ddAMP to the 3' end of GΔ228-Uve1p-generated DNA cleavage products and indicates the presence of a 3'-hydroxyl terminus. These results show that the 3' and 5' termini of the products of GΔ228-Uve1p-mediated cleavage of substrates containing single base mismatches are identical to those generated following cleavage of substrates containing CPDs or 6-4PPs.

To verify that the Uve1p mismatch endonuclease activity observed was not the result of trace endonucleolytic contamination from the *S. cerevisiae* expression system and to determine whether full length Uve1p was also capable of mismatch endonuclease activity, extracts from cells overexpressing GFL-Uve1p, GΔ228-Uve1p, and GST tag alone were tested for their abilities to cleave 5' end-labeled *CX/AY-31mer. Both GFL-Uve1p and GΔ228-Uve1p cleaved the base mismatch-containing substrate at positions 0, -1, and -2 (Fig. 11B). We also observed a weak 3' to 5' exonucleolytic activity associated with both crude GFL-Uve1p preparations and purified GΔ228-Uve1p which shortened the Uve1p-mediated cleavage products by one to three nucleotides (Fig. 11B, lanes 1 and 2). These shorter products are not due to additional cleavages by Uve1p mismatch endonuclease activity because they are not observed in identical experiments with 3' end-labeled substrates. Purified Δ228-Uve1p obtained following thrombin cleavage of the GST tag also possessed mismatch endonuclease activity. In contrast, no cleavage of mismatch-containing substrates was observed when extracts from cells transfected with vector expressing only the GST tag were tested. Thus, both GFL-Uve1p and its more stable, truncated version, GΔ228-Uve1p, both possess mismatch endonuclease activities.

GΔ228-Uve1p mismatch endonuclease and GΔ228-Uve1p UV photoproduct endonuclease share similar properties and compete for the same substrates. GΔ228-Uve1p requires divalent cations for activity and exhibits optimal activity against UV photoproducts in the presence of 10 mM $MgCl_2$ and 1 mM $MnCl_2$. Omission of divalent cations from the reaction buffer abolished GΔ228-Uve1p mismatch endonuclease activity on 5' end-labeled *CS/AY-31mer. The pH optimum for GΔ228-Uve1p mismatch endonuclease activity on this same substrate was found to be 6.5, which corresponds to the pH where optimal activity is observed against UV photoproducts.

To further confirm that the mismatch endonuclease activity was mediated by GΔ228-Uve1p, a substrate competition experiment was performed with CPD-30mer, a known Uve1p substrate which contains a centrally located UV photoproduct (CPD). Addition of increasing amounts of unlabeled CPD-30mer resulted in a significant, concentration-dependent decrease in GΔ228-Uve1p-mediated mismatch endonuclease activity against 3' end-labeled *CX/AY31mer (C/A mispair) (Fig. 12). In contrast, increasing amounts of the undamaged oligo GX/CY-31mer (G/C base pair) had only a modest inhibitory effect, and inhibition did not increase with increasing amounts of added oligo, indicating a non-specific binding to Uve1p within this concentration range. In a similar experiment both unlabeled CPD-30mer and CX/AY-31mer (C/A mispair) were more potent inhibitors of 3' end-labeled *CX/AY-31mer cleavage compared to unlabeled GX/CY-31mer. The effective competition by CPD-30mer for mismatch endonuclease activity indicates that both base mismatch and UV photoproduct endonuclease activities are associated with GΔ228-Uve1p.

Uve1p incises only one strand of a duplex containing a base mismatch. Since Uve1p recognizes all possible base mismatch combinations, we determined whether the enzyme could incise both strands on the same molecule resulting in a DNA double strand break. An oligonucleotide (*CX/AY-41mer) was designed such that the base mispair was placed in the center of the oligonucleotide. GΔ228-Uve1p was incubated with 3' end-labeled *CS/AY-41mer under standard conditions, and the DNA strand scission products were analyzed on both non-denaturing and denaturing gels (Fig. 13A-13B). In the event that GΔ228-Uve1p created a DNA double strand break by incising 5' to the base mismatch site on the two

complementary strands, the resulting products would possess an electrophoretic mobility similar to those created by the restriction enzyme DdeI (which cleaves adjacent to the mismatch) when analyzed on a non-denaturing polyacrylamide gel. In contrast, if GΔ228-Uve1p incises on either (but not both) complementary strands, then the resulting product would be a full-length duplex containing a single strand nick which would co-migrate with uncut duplex *CX/AY-41mer on a non-denaturing gel. Non-denaturing gel analysis of GΔ228-Uve1p-treated *CX/AY-41mer generated a product with an electrophoretic mobility identical to the untreated duplex with no products detected corresponding to those created by a double strand break (Fig. 13A). Denaturing gel analysis revealed a GΔ228-Uve1p-generated DNA strand scission product resulting from a single strand break of the labeled strand of either *CX/AY-41mer or CX/*GY-41mer. Together with the non-denaturing gel analysis, these results indicate that within the GΔ228-Uve1p substrate population, nicks occur on one or the other, but not both strands (Fig. 13B). These results show that GΔ228-Uve1p nicks only one of the two strands containing a base mismatch and that it does not make double strand breaks in duplex DNA. Similarly, double strand breaks are not made in DNA molecules containing other structural distortions.

Without wishing to be bound by theory, it is believed that GΔ228-Uve1p possesses strand specificity directed towards the 3' terminus. Mismatched bases in duplex DNA are distinct from damaged DNA in the sense that both of the bases are usually undamaged per se, yet one is an inappropriate change in the nucleotide sequence and must be identified as such and removed. If Uve1p participates in MMR *in vivo*, how might it distinguish between the correct and incorrect bases in a mispair? One possibility is that proximity of the mispaired base to either the 3' or 5' terminus targets Uve1p mismatch endonuclease activity to a particular strand. For example, in DNA synthesis, chain growth proceeds from the 5' to the 3' terminus and newly-generated base misincorporations on the synthesized strand would be located in close proximity to the 3' terminus. Initiating the removal of such bases by a mismatch repair protein might involve association with a region of DNA in the vicinity of the 3' terminus, followed by targeting of the mispaired base located on that strand. To investigate this possibility, a series of 3' end-labeled oligonucleotides were generated that contained a C/A mispair located at various distances from the ends (Table 1B). The ability of GΔ228-

Uve1p to incise the C-containing strand as a function of the distance of C (of the C/A mispair) from the 3' terminus was assessed by quantifying the GΔ228-Uve1p mismatch endonuclease-generated DNA stand scission products following denaturing gel analysis. A minimum level of mismatch cleavage was observed for C at a distance of 16 bp from the 3' terminus and gradually increased to a maximum for C at a distance of 16 bp from the 3' terminus. Closer placement (11 bp) of C to the 3' terminus resulted in a decrease in mismatch endonuclease activity with a complete loss of activity observed at a distance 6 bp from the 3' terminus. The mismatched base located on the strand in closest proximity to the 3' terminus is cleaved preferentially by GΔ228-Uve1p.

uve1 null mutants exhibit a mutator phenotype. We have examined the spontaneous mutation rate of *uve1::ura4*$^+$ disruption mutants as assayed by the ability to form colonies resistant to the toxic arginine analog L-canavanine. Uptake of L-canavanine in *S. pombe* is mediated by an arginine permease encoded by the *can1*$^+$ gene (Fantes, P. and Creanor, J. [1984] *J. Gen. Microbiol.* **130**:3265-3273). Mutations in *can1*$^+$ eliminate the uptake of L-canavanine, and mutant cells are able to form colonies on medium supplemented with L-canavanine, whereas wild type cells cannot. We have compared the rate of spontaneous mutagenesis at the *can1*$^+$ locus in *uve1::ura4*$^+$ disruption mutants (Sp362) to both a negative control (wild type, 972) and a positive control, *pms1::ura4*$^+$ (see Example 11 hereinbelow). The *pms1* gene product is a homolog of *E. coli* MutL, and loss of *pms1* causes a strong mitotic mutator phenotype and increased postmeiotic segregation (Schar et al. [1997] *Genetics* **146**:1275-1286).

To determine the relative sensitivity of each yeast strain to L-canavanine, 200 cells from mid-log phase cultures were plated onto PMALU$^-$ plates supplemented with increasing concentrations of L-canavanine. Each of the strains was equally sensitive to L-canavanine. All strains were viable in the presence of lower concentrations of L-canavanine up to and including concentrations of 2.2 μg/ml, while concentrations higher than this were toxic to all strains. However, the colonies which grew in the presence of 2.2 μg/ml L-canavanine were smaller in diameter than the colonies which grew in the presence of lower concentrations.

The mean spontaneous mutation rate of each of the three strains was examined using fluctuation analyses. Single colonies grown on PMALU^g plates were used to inoculate liquid PMALU^g cultures which were grown to saturation. 10^7 cells were plated onto PMALU^g containing 75 µg/ml L-canavanine sulfate. The number of colonies on 24 plates for each strain was counted after 8 days incubation at 30°C. Both *uve1::ura4*$^+$ and *pms1::ura4*$^+$ strains showed an elevated number of resistant colonies compared to wild type. Additionally, the range of values for *uve1::ura4*$^+$ was broader and higher than for either wild type or *pms1::ura4*$^+$ and included two confluent plates scored as containing >5000 colonies. the mean rate of mutation was estimated using the method of the median (Lea and Coluson [1943] *J. Genet.* **49**:264-284) using the median values. The calculated mutation rates are 1.5 x 10^{-7} (wild type), 9.7 x 10^{-7} (*uve1::ura4*$^+$), and 2.0 x 10^{-6} (*pms1::ura4*$^+$), indicating that *uve1::ura4*$^+$ mutants have a spontaneous mutation rate approximately 6.5-fold greater than wild type and 2-fold lower than *pms1::ura4*$^+$. See Table 4 for a summary of results. Thus, loss of Uve1p confers a spontaneous mutator phenotype in *S. pombe*. In the mutation fluctuation analysis, a wide range of mutant colonies was observed for *uve1::ura4*$^+$ compared to *uve1::ura4*$^+$, suggesting that the pathways leading to mutation due to elimination of *uve1* and *pms1* are likely to be mechanistically different.

The finding that Uve1p recognizes all potential DNA base mispair combinations indicates that, in addition to its UV photoproduct cleavage activity, it is a diverse mismatch endonuclease with broad substrate specificity. In this regard, Uve1p is similar to *E. coli* endonuclease V (Yao, M. and Kow, Y.W. [1994] *J. Biol. Chem.* **269**:31390-31396), a *S. cerevisiae* and human "all-type" mismatch endonuclease (Chang, D.Y. and Lu, A.L. [1991] *Nucl. Acids Res.* **19**:4761-4766; Yeh et al. [1991] *J. Biol. Chem.* **266**:6480-6484) and calf thymus topoisomerase I (Yeh et al. [1994] *J. Biol. Chem.* **269**:15498-15504) which also recognize all potential base mismatch combinations. These enzymes incise DNA at each of the twelve base mispairs with variable efficiencies and either to the 5' (human all-type mismatch endonuclease) or 3' (*E. coli* endonuclease V) sides of a mismatch. Uve1p shows a preference for *C/C and *C/A mispairs, a property similar to the human all-type mismatch endonuclease (Yeh et al. [1991] *supra*). In contrast, the strong preference of Uve1p for *G/G

mispairs is a property which distinguishes Uve1p from all other mismatch endonucleases identified to date.

The biochemical properties of Uve1p-mediated mismatch cleavage and the spontaneous mutator phenotype displayed by *uve1* null mutants suggest that Uve1p is involved in MMR *in vivo*. The preference for making incisions on the strand harboring the mispaired base nearest to the 3' terminus reflects a discrimination strategy that might specifically target newly misincorporated bases during replication. Uve1p-generated incision 5' to the base mismatch site could be followed by a 5' to 3' exonuclease activity such as that mediated by *S. pombe* exonuclease I (Szankasi, P. and Smith, G.R. [1995] *Science* **267**:1166-1169) or the FEN-1 homolog Rad2p (Alleva, J.L. and Doetsch, P.W. [1998] *Nucl. Acids Res.* **26**:3645-3650) followed by resynthesis and ligation.

S. pombe possesses at least two distinct mismatch repair systems and whether Uve1p mediates a role in either of these or represents a third, novel pathway is not known at present. The proposed major pathway does not recognize C/C mismatches and has relatively long (approximately 100 nt) repair tracts (Schar, P. and Kohli, J. [1993] *Genetics* **133**:825-835). Uve1p is thought to participate in a relatively short patch repair process which utilizes Rad2p (a FEN-1 homolog) DNA polymerase δ, DNA ligase and accessory factors (Alleva et al. [1998] *Nucl. Acids Res.* **26**:3645-3650). Based on these properties, it is unlikely that Uve1p is involved in a long tract mismatch repair system. The second, presumably less frequently utilized, (alternative) pathway recognizes all potential base mismatch combinations and has a repair tract length of about 10 nucleotides (Schar and Kohli [1993] *supra*). These features of the alternative mismatch repair pathway are consistent with the repair properties of Uve1p based on recognition of C/C mismatches and short repair patch.

Unlike in repair of UV photoproducts, it is not clear in mismatch repair which base represents the nucleotide that needs to be removed. This can be explained by our finding that Uve1p prefers a mispaired base located near the 3' terminus of a duplex, which is consistent with Uve1p mediating mismatch repair for either leading or lagging strand synthesis during DNA replication. The preference for making incisions on the strand of the base nearest to the

3' terminus suggests a discrimination strategy to specifically target newly synthesized misincorporated bases. On the other hand, G/G, C/C mismatches are not frequently occurring base misincorporations encountered during replication, although they are among the most efficiently cleaved by Uve1p. A second role for Uve1p is in the correction of mismatched bases formed as a result of homologous recombination events where G/G and C/C mismatches would be expected to occur. A third role for Uve1p is in the repair of base bulges and loops generated as a result of primer-template misalignments during replication. Preliminary studies show that Uve1p mediates strand cleavage 5' to small bulges.

What is the structural basis for lesion recognition by Uve1p? Previous studies with Uve1p have focused exclusively on its role in the repair of UV light-induced DNA damage, resulting in the notion that this enzyme functions in the repair of UV photoproducts exclusively, hence the prior name UVDE (UV damage endonuclease), now Uve1p. The results of this study clearly indicate a much broader involvement of Uve1p in *S. pombe* DNA repair and show that many other types of DNA lesions are recognized by this versatile repair protein. For example, we have recently found that Uve1p recognizes and incises DNA substrates containing uracil, dihydrouracil, cisplatin-induced adducts as well as small base bulges. The molecular basis for substrate recognition by Uve1p is not obvious, but without wishing to be bound by theory, it is believed to be due in part to disruption of normal Watson-Crick base pairing and the corresponding changes expected in the electronic characteristics of the major and minor grooves of B-DNA.

Besides initiating repair of DNA containing UV damage including CPDs and 6-4PPa, UVDE and the truncated UVDE polypeptide of the present invention (Δ228-UVDE and/or GST-Δ228-UVDE) also initiate repair via cleavage of DNA duplexes containing the following base pair mismatches: C/A; G/A; G/G; A/A; and C/T. These experiments were conducted with GST-Δ228-UVDE. We also confirmed that the C/A mismatch is cleaved by Δ228-UVDE; it should also recognize the others. In addition, both GST-Δ228-UVDE and Δ228-UVDE recognize and cleave an oligonucleotides containing a GG-platinum diadduct formed by the antitumor agent cis-dichlorodiammineplatinum (II) (also known as cisplatin). Thus the substrate specificity range for UVDE is much broader than originally thought.

Recognition of the truncated UVDE polypeptide to initiate mismatch repair was made possible due to the increased stability of the presently exemplified truncated UVDE polypeptide in substantially purified form.

Skin cancers associated with sunlight exposure are the most common worldwide human cancers. The primary DNA damage from exposure to sunlight are 6-4 PPs and CPDs. Since UVDE can augment cells defective in DNA repair, the stable truncated UVDE fragments of the present invention will be valuable therapeutic agents for correcting DNA repair defects in sunlight-sensitive and skin cancer-prone individuals, for example individuals with the genetic disease xeroderma pigmentosum. Additionally, GST-Δ228-UVDE and Δ228-UVDE can be used as protective agents against sunlight-induced skin damage in normal individuals because they can augment the existing DNA repair levels of CPDs and 6-4 Pps and other DNA damage.

Homologs of the *S. pombe* UVDE protein have been identified by BLAST searching of sequence database (Genbank, TIGR) using the UVDE amino acid sequence: *N. crassa* (Genbank Accession No. BAA 74539), *B. subtilis* (Genbank Accession No. 249782), human (Genbank Accession No. AF 114784.1, methyl-CpG binding endonuclease) and a *Deinococcus radiodurans* sequence located from the TIGR database. The amino acid sequences of these proteins are given in SEQ ID NO:36 (*N. crassa*), SEQ ID NO:37 (*B. subtilis*), SEQ ID NO:38 (*Homo sapiens*) and SEQ ID NO:39 *(D. radiodurans*). The *D. radiodurans* coding sequence can be generated using the genetic code and codon choice according to the recombinant host in which the protein is to be expressed, or the natural coding sequence can be found on the TIGR database, *D. radiodurans* genomic sequence in the region between bp 54823 and 60981.

The regions of the *S. pombe* UVDE protein which are most conserved in the foregoing homologs are amino acids 474-489, 535-553, 578-611, 648-667, 711-737 and 759-775 of SEQ ID NO:2.

The stable truncated UVDE derivatives of the present invention are useful to treat or prevent diseases caused by cyclobutane pyrimidine dimers or (6-4) photoproducts or DNA mismatch, abasic sites or other distortions in the structure of double stranded DNA through the application of skin creams which can deliver GST-Δ228-UVDE and Δ228-UVDE to the appropriate living cells or via other routes of administration with compositions suitable for the route of administration, as is well understood in the pharmaceutical formulation art. GST-Δ228-UVDE or Δ228-UVDE can be incorporated into liposomes, and the liposomes can be applied to the surface of the skin, whereby the encapsulated GST-Δ228-UVDE and Δ228-UVDE products traverse the skin's stratum corneum outer membrane and are delivered into the interior of living skin cells. Liposomes can be prepared using techniques known to those skilled in the art. A preferred liposome is a liposome which is pH sensitive (facilitates uptake into cells). Preparation of pH sensitive liposomes is described in U.S. Pat. No. 5,643,599, issued to Kyung-Dall et al.; and 4,925,661 issued to Huang. The GST-Δ228-UVDE and Δ228-UVDE polypeptides can be entrapped within the liposomes using any of the procedures well known to those skilled in the art. See, e.g., the Examples and U.S. Pat. Nos. 4,863,874 issued to Wassef et al.; 4,921,757 issued to Wheatley et al.; 5,225,212 issued to Martin et al.; and/or 5,190,762 issued to Yarosh.

The concentration of liposomes necessary for topical administration can be determined by measuring the biological effect of GST-Δ228-UVDE and Δ228-UVDE, encapsulated in liposomes, on cultured target skin cells. Once inside the skin cell, GST-Δ228-UVDE or Δ228-UVDE repairs CPDs or 6-4 Pps in damaged DNA molecules and increases cell survival of those cells damaged by exposure to ultraviolet light.

Polyclonal or monoclonal antibodies specific to GST-Δ228-UVDE and Δ228-UVDE allow the quantitation of GST-Δ228-UVDE and Δ228-UVDE entrapped into liposomes. GST-Δ228-UVDE and Δ228-UVDE antibodies also allow tracing of the truncated UVDE polypeptides into skin cells.

Standard techniques for cloning, DNA isolation, amplification and purification, for enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and

the like, and various separation techniques are those known and commonly employed by those skilled in the art. A number of standard techniques are described in Sambrook et al. (1989) *Molecular Cloning*, Second Edition. Cold Spring Harbor Laboratory, Plainview, New York; Maniatis et al. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Plainview, New York; Wu (ed.) (1993) *Meth. Enzymol.* Part I; Wu (ed.) (1979) *Meth Enzymol.* 65; Miller (ed.) (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York; Old Primrose (1981) *Principles of Gene Manipulation*, University of California Press, Berkeley; Schleif and Wensink (1982) *Practical Methods in Molecular Biology*; Glover (ed.) (1985) *Nucleic Acid Hybridization*, IRL Press, Oxford, UK; and Setlow and Hollaender (1979) *Genetic Engineering*: *Principles and Methods*, Vols. 1-4, Plenum Press, New York. Abbreviations and nomenclature, where employed, are deemed standard in the field and commonly used in professional journals such as those cited herein.

Each reference cited in the present application is incorporated by reference herein to the extent that it is not inconsistent with the present disclosure.

The following examples are provided for illustrative purposes, and are not intended to limit the scope of the invention as claimed herein. Any variations in the exemplified articles which occur to the skilled artisan are intended to fall within the scope of the present invention.

EXAMPLES

Example 1. Strains, enzymes, plasmids and genes.

E. coli Top10 (Invitrogen Corp., San Diego, CA) was used for subcloning and plasmid propagation. *S. cerevisiae* strain DY150 used for protein expression and the *S. cerevisiae* expression vector pYEX4T-1 were purchased from Clontech (Palo Alto, CA).

S. pombe strains used in this study include 972, h^{-s} (Leupold, U. [1970] *Meth. Cell Physiol.* **4**:169-177); PRS301, h^{-s} *pms1::ura4*$^{+}$ (Schar et al. [1993] *Genetics* **146**:1275-1286); SP30, h^{-s} *ade6-210 leu-32 ura4-D18* (Davey et al. [1998] *Mol. Cell. Biol.* **18**:2721-2728).

Sp362 (*h*^{s} *ade6-210 leu1-32 ura4-D18 uve1::ura4*$^{+}$) was constructed by transforming Sp30 with a linearized, genomic *uve1*$^{+}$ fragment derived from pgUV2 (Davey et al. [1997] *Nucl. Acids Res.* **25**:1005-1008) in which nucleotides 215 (*EcoRI*) to 1045 (*ClaI*) of *uve1*$^{+}$ were replaced with the *ura4+* gene. Extracts of Sp362 contained no detectable Uve1p activity against CPD-30mer. Cultures were grown in pombe minimal medium (PM) (Leupold [1970 *supra*) with 3.75 g/l glutamate replacing ammonium chloride as the nitrogen source (Fantes, P. and Creanor, J. [1984] *J. Gen. Microbiol.* **130**:3265-3273), and were supplemented with 150 mg/l of each adenine, leucine and uracil (PMALU^{g}). Solid media was prepared by addition of 20 g/l agar. L-canavanine sulfate was sterilized prior to addition to the medium.

Purified mismatch repair endonuclease, *E. coli* endonuclease V (Yao, M. and Kow, Y.W. [1997] *J. Biol. Chem.* **272**:30774-30779) was a gift from Yoke Wah Kow (Atlanta, GA).

Example 2. Amplification of the *uvde* (*uve1*) gene from *S. pombe*.

A cDNA library purchased from ATCC was amplified by PCR, using the sense primer: 5'-TGAGGATCCAATCGTTTTCATTTTTTAATGCTTAGG-3' (SEQ ID NO:9) and the antisense primer: 5'-GGCCATGGTTATTTTTCATCCTC-3' (SEQ ID NO:10). The gene fragment of interest was amplified in the following manner. Four hundred nanograms of template DNA (*S. pombe* cDNA library) was incubated with the upstream and downstream primers (300nM) in the presence of Pwo DNA polymerase (Boehringer Mannheim, Indianapolis, IN) in 10 mM Tris-HCl (pH 8.85), 25 mM KCl, 5 mM $(NH_4)_2SO_4$, 2mM $MgSO_4$ and 200 μM of dNTPs. The DNA was initially denatured at 94°C for 2 min. Three cycles of denaturation at 94°C for 15 sec, annealing at 45°C for 30 sec and primer extension at 72°C for 2 min were followed by twenty cycles using 50°C as the annealing temperature. All other incubation times and temperatures remained the same. The amplification was completed by a final primer extension at 72°C for 7 min.

Example 3. Amplification of the Δ228-UVDE gene-encoding fragment from *S. pombe*.

PCR was used to produce a truncated DNA fragment of the full-length *S. pombe uvde* gene which encodes a protein product containing a deletion of 228aa from the N-terminal

portion of the full-length *S. pombe* UVDE protein. The following primers were used in the PCR reaction to amplify the gene fragment which encodes Δ228-UVDE: sense primer 5'-AATGGGATCCGATGATCATGCTCCACGA-3' (SEQ ID NO:11) and the antisense primer 5'-GGGATCCTTATTTTTCATCCTCTTCTAC-3' (SEQ ID NO:12). PCR conditions were as described in Example 2.

Example 4. Purification of Δ228-UVDE and full-length UVDE.

The amplified UVDE gene coding fragments were cloned into the *BamHI* and *SmaI* restriction sites of pYEX 4T-1. The Δ228-UVDE gene coding fragments were cloned into the *BamHI* restriction site of pYEX4T-1 (Clontech, Palo Alto, CA). In the pYEX4T-1 vector, the coding region of both the proteins is expressed in frame with a glutathione-S-transferase (GST) leader sequence to generate a fusion protein of GST linked to the N-terminus of UVDE which is under the control of the CUP1 promoter (Ward et al., 1994). The subcloned plasmids were checked for orientation by restriction analysis and were then transformed into *S. cerevisiae*, DY150 cells, using the alkali cation method (Ito et al. [1983] *supra*). A single positive clone was picked and grown at 30° C until mid log phase. Cultures in mid log phase were induced with 0.5 mM $CuSO_4$. Cells (500 mL) were harvested 2 hr after induction and lysed with glass beads in 50 mM Tris (pH 7.5), 100 mM EDTA, 50 mM NaCl, 10 mM β-mercaptoethanol, 5% glycerol in the presence of 10 ng/mL pepstatin, 3 nN leupeptin, 14.5 mM benzamidine, and 0.4 mg/mL aprotinin. The cell lysate was then dialyzed overnight in buffer minus EDTA. The whole cell homogenate was separated into soluble and insoluble fractions by centrifugation at 45,000 X g for 20 min. The soluble proteins (120 mg) were applied to a 2 mL glutathione-Sepharose-affinity column (Pharmacia, Piscataway, NJ). All purification steps were carried out at 4° C and are similar to the strategies employed for the purification of other types of GST-tagged proteins (Ward, A.C. et al. [1994] *Yeast* **10**:441-449; Harper, S. and Speicher, D. [1997] in *Current Protocols in Protein Sci.* [Coligan, J. et al., Eds) pp. 6.6.1-6.6.21, John and Wiley & Sons). Unbound proteins were removed by washing with 30 mL phosphate-buffered saline (pH 7.4), 5 mM EDTA, 0.15 mM PMSF. GST-Δ228-UVDE was eluted (100-200 μL fractions) with 10 mM glutathione in 50 mM Tris (pH 7.4) or cleaved on the column with excess of thrombin as previously described (Harper and Speicher, 1997) to generate Δ228-UVDE without the GST tag. SDS-PAGE analysis of

flow-through, wash, elution, and thrombin cleavage fractions indicated the extent of purification or GST tag removal via thrombin cleavage (Fig. 1A-1B).

Example 5. GST preparation.

S. cerevisiae (DY150) cells were transformed with the pYex4T-1 expression vector without any insert (i.e., expressing gluthathione-S-transferase [GST] alone). These cultures were induced with $CuSO_4$ and cell lysates were prepared as described for the Uve1p proteins. Purified recombinant GST was affinity-purified on a gluthathione sepharose column in an identical manner to GΔ228-Uve1p (see above) and was included in all of the assays performed in this study as a control for trace amounts of potential contaminating endonucleases in the Uve1p protein preparations.

Example 6. UVDE activity assay and optimization of reaction conditions.

Crude and purified full-length UVDE, GST-Δ228-UVDE and Δ228-UVDE were tested for activity on an oligodeoxynucleotide substrate (CPD-30mer) containing a single cis-syn cyclobutane pyrimidine dimer embedded near the center of the sequence. The sequence of the CPD-containing strand is: 5'-CATGCCTGCACGAAT^TAAGCAATTCGTAAT-3' (SEQ ID NO:13) . The CPD-containing DNA molecule was synthesized as described by Smith, C.A. and Taylor, J.S. (1993) *J. Biol. Chem.* **268**:11143-11151. The CPD-30mer was 5' end labeled with [γ-^{32}P]ATP (Amersham, 3000 Ci/mmol) using polynucleotide kinase (Tabor, 1989). For UVDE reactions with end labeled CPD-30mer, approximately 10 fmol of 5' end labeled CPD 30-mer was incubated with 5-100 ng of Δ228-UVDE or GST-Δ228-UVDE, in 200 mM Hepes (pH 6.5), 10 mM $MgCl_2$, 1 mM $MnCl_2$, 150 mM NaCl for 15 min at 37° C 10-20 μL reaction volume). The reaction products were analyzed on 20% denaturing (7 M urea) polyacrylamide gels (DNA sequencing gels) as previously described (Doetsch, et al., 1985). The DNA species corresponding to the uncleaved CPD-30mer and cleavage product (14-mer) were analyzed and quantified by phosphorimager analysis (Molecular Dynamics Model 445SI) and autoradiography.

In other experiments, reactions with various Uve1p preparations were carried out in a total volume of 20 μL, and contained reaction buffer (20 mM Hepes, pH 6.5, 100 mM NaCl,

10 mM $MgCl_2$ and 1 mM $MnCl_2$) and end-labeled oligonucleotide substrate (10-30 fmol). The substrate/buffer mix was incubated for 20 min at 37°C with Uve1p. In the case of G-Uve1p and GΔ228-Uve1p, crude cell lysates (5 μg of protein) were used for all assays. Fifty ng of affinity-purified GΔ228-Uve1p (0.75 pmol) and Δ228-Uve1p (1.2 pmol) were incubated with all of the UV-induced photoproducts. For all other assays 2 μg of affinity-purified GΔ228-Uve1p (30 pmol) and Δ228-Uve1p (48 pmol) were incubated with the substrates. Two μg of affinity-purified recombinant GST (72 pmol) was incubated with each substrate under Δ228-Uve1p optimum reaction conditions to control for potential contaminating nuclease activities which may be present in the Uve1p preparations and to determine the specificity of the Uve1p cleavage reaction. DNA repair proteins (*E. coli* exonuclease III, *E. coli* endonucleases III and IV, *E. coli* uracil DNA glycosylase and *S. cerevisiae* endonuclease III-like glycosylase [Ntg]) specific for each oligonucleotide substrate were also incubated with these substrates under their individual optimum reaction conditions, as a means to determine the specific DNA cleavage sites of Uve1p. The reaction products were analyzed on 20% denaturing (7M urea) polyacrylamide gels (DNA sequencing-type gels) as described previously (Doetsch et al. [1985] *Nucl. Acids Res.* **13**:3285-3304). The DNA bands corresponding to the cleaved and uncleaved substrate were analyzed and quantified by phosphorimager analysis (Molecular Dynamics Model 445SI) and autoradiography.

Example 7. Oligonucleotides containing DNA damage.

The DNA damage-containing oligonucleotides used as substrates in this study are presented in Table 1A. The structure of each damaged lesion is presented in Figure 1. The 30-mer cs-CPD-containing oligonucleotide (cs-CPD-30mer) was prepared as described previously (Smith, C.A. [1993] *J. Biol. Chem.* **268**:11143-11151). The 49-mer oligonucleotides containing a cs-CPD (cs-CPD-49mer), a ts I-CPD (tsI-CPD-49mer), a ts II-CPD (tsII-CPD-49mer), a 6-4PP (6-4PP-49mer) and a Dewar isomer (Dewar-49mer) were synthesized as described previously (Smith, C.A. and Taylor, J-S. [1993] *J. Biol. Chem.* **268**:11143-11151). The oligonucleotide containing a platinum-DNA GG diadduct (Pt-GG-32mer) and its complementary strand were prepared as previously described (Naser et al. [1988] *Biochemistry* **27**:4357-4367). The uracil-containing oligonucleotide (U-37mer), the

undamaged oligonucleotides and the complementary strand oligonucleotides for all the substrates were synthesized by the Emory University Microchemical Facility. The DHU-containing oligonucleotide (DHU-37mer) was synthesized by Research Genetics (Birmingham, AL). The oligonucleotides containing inosine (I-31mer) and xanthine (Xn-31mer) and their complementary strands were a gift from Dr. Yoke Wah Kow (Emory University, Atlanta, GA). The 8-oxoguanine-containing 37-mer (8-oxoG-37mer) was synthesized by National Biosciences Inc. (Plymouth, MN).

Labeled oligonucleotide substrates were prepared as follows: The cs-CPD-30mer, the 49mer UV photodamage-containing oligonucleotides and the Pt-GG-32mer were 5' end-labeled with [γ-^{32}P] ATP (Amersham, 3000 Ci/mmol) using polynucleotide kinase (Tabor, S. [1985] in *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley [Interscience], New York, NY). the oligonucleotides U-37mer, DHU-37mer, I-31mer, Xn-31mer and 8-oxoG-37mer were 3' end-labeled using terminal transferase and [α^{32}P] ddATP (Amersham, 3000 Ci/mmol) (Tu, C. and Cohen, S.N. [1980] *Gene* **10**:177-183). End-labeled duplex oligonucleotides were gel-purified on a 20% non-denaturing polyacrylamide gel. The DNA was resuspended in ddH_2O and stored at -20 C.

The AP substrate was prepared as described hereinbelow. 5' end-labeled, duplex U-37mer (20-50 pmol) was incubated with uracil DNA glycosylase (UDG, 6 units) for 30 minutes at 37°C in UDG buffer (30 mM Hepes-KOH, pH 7.5, 1 mM EDTA, and 50 mM NaCl) to generate the AP site-containing oligonucleotide (AP-37mer). The DNA was extracted with PCIA (phenol-chloroform-isoamylalcohol, 29:19:1, v/v/v) equilibrated with HE buffer (10 mM Hepes-KOH pH 8.0, 2 mM EDTA) with 0.1% 8-hydroxyquinoline, and was evaluated for its AP site content by cleavage with 0.1 M piperidine at 90°C for 20 minutes.

The CPD-30mer Uve1p substrate (see herein and Kaur et al. [1998] *Biochemistry* **37**:11599-11604) containing a centrally embedded, cis-syn TT cyclobutane pyrimidine dimer was a gift from John-Stephen Taylor (St. Louis, MO). All other oligonucleotide substrates (Table 1) for mismatch endonuclease experiments were synthesized by Operon, Inc.

(Alameda, CA) or IDT, Inc. (Coralville, IA). All oligonucleotides were gel-purified and subjected to DNA sequence analysis for sequence confirmation. Oligonucleotides were 5' end-labeled with polynucleotide kinase using 50 μCi [γ-^{32}P] ATP (Amersham, 3000 Ci/mmol) as previously described (Bowman et al. [1994] *Nucl. Acids Res.* **22**:3026-3032). 3' end-labeled oligonucleotides were prepared by incubating 10 pmol of the indicated oligonucleotide with 10 units of terminal deoxynucleotidyl transferase (TdT, Promega) and 50 μCI of [α-^{32}P] ddATP (Amersham, 3000 Ci/mmol) as previously described (Bowman et al. [1994] *supra*).

Example 8. Establishment of optimal reaction conditions.

The optimal reaction conditions for UVDE cleavage of CPD-30mer were established by varying the NaCl concentration, divalent cation ($MnCl_2$, and $MgCl_2$) concentration, or by varying the pH of the reaction buffer in the reaction. The buffers (20 mM at the indicated pH range) were as follows: sodium citrate (pH 3-6), Hepes-KOH (pH 6.5-8), and sodium carbonate (pH 9-10.6). The optimum temperature required for enzyme activity was determined by pre-incubating the enzyme and the substrate in the reaction buffer at a specific temperature for 10 min prior to mixing UVDE and CPD-30mer. The reaction was stopped by phenol-chloroform-isoamyl alcohol extraction and the reaction products were analyzed on DNA sequencing gels as described above. From these experiments the following standard reaction conditions were established: 20 mM Hepes (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$, 1 mM $MnCl_2$, 30°C or at 37°C for 20 minutes.

Example 9. Kinetic assays

Enzyme reactions were carried out with 5 nM Δ228-UVDE or 11.5 nM GST-Δ228-UVDE in 20 mM Hepes (pH 6.5) in 10 mM $MgCl_2$, 1 mM $MnCl_2$, 100 mM NaCl. 5' End labeled CPD-30mer concentrations were varied from 25-250 nM in a final reaction volume of 15 μL for 0-3 minutes at 37° C. Initial enzyme velocities (V_i) were measured for each substrate concentration as nM of product formed per second. The apparent K_m, V_{max}, and turnover number (K_{cat}) were determined from Lineweaver Burk plots of averaged data (± standard deviations) from three independent experiments.

Example 10. Analysis of Uve1p Mismatch Repair Activity.

Reactions with GΔ228-Uve1p were carried out by incubating approximately 100 fmol of labeled oligonucleotide substrate with 100-150 ng of purified GΔ228-Uve1p in 20 mM Hepes (pH 6.5), 10 mM $MgCl_2$, 1 mM $MnCl_2$, and 150 mM NaCl for 20 minutes at 37°C (10-20 μl final volume). Reactions with crude preparations of GFL-Uve1p were carried out with 20-30 μg of cell extract incubated with the appropriate substrate in 20 mM Hepes (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$ and 1 mM $MnCl_2$ at 37°C for 20 minutes. The reaction products were processed by extracting with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1), ethanol-precipitation, resuspension and analysis on 20% denaturing (7 M urea) polyacrylamide (DNA sequencing) gels as previously described (Kaur et al. [1998] *supra*). The DNA species corresponding to the uncleaved substrate and Uve1p-mediated DNA strand scission products were analyzed and quantified by phosphorimager analysis (Molecular Dynamics model 445SI) and autoradiography.

Terminal analysis of the mismatch cleavage products was carried out as follows. GΔ228-Uve1p was incubated with 3' end-labeled *CX/AY-31mer under standard reaction conditions at 37°C for 20 minutes. The ethanol-precipitated reaction products were incubated with either 10 units of calf intestinal phosphatase (CIP, Promega, Madison WI) at 37°C for 30 minutes or with 10 units of T4 polynucleotide kinase (PNK, New England Biolabs) and 50 pmol ATP as previously described (Bowman et al. [1994] *supra*). The reaction products were analyzed on 20% denaturing polyacrylamide gels as described above for Uve1p activity assays. Differences in electrophoretic mobilities of kinase-treated versus untreated DNA strand scission products indicated the presence or absence of a pre-existing 5'-phosphoryl group (Bowman et al. [1994] *supra*).

3' terminal analysis of the mismatch cleavage products was carried out as follows. To determine the chemical nature of the 3' terminus of GSTΔ228-Uve1p-mediated DNA strand scission products, 5' end-labeled *CX/AY-31mer was incubated with GΔ228-Uve1p as described above. The ethanol-precipitated, resuspended reaction products were then treated with 10 units of TdT and ddATP as previously described (Bowman et al [1994] *supra*).

Samples were processed and analyzed on polyacrylamide gels as described above for 5' terminal analysis.

To determine the pH optimum for Uve1p-mediated mismatch cleavage, 100 fmol of 3' end-labeled *CX/AY-31mer was incubated with approximately 100 ng of GΔ228-Uve1p with 10 mM $MgCl_2$ and 1 mM $MnCl_2$ in 20 mM reaction buffer of different pH ranges (pH 3.0-10.6). The buffers were as follows: sodium citrate (pH 3.0-6.0), Hepes-KOH (pH 6.5-8.0), and sodium carbonate (pH 9.0-10.6). The reaction products were analyzed on a 20% denaturing polyacrylamide gel and the optimal pH was calculated as previously described for Uve1p cleavage of CPD-30mer (Kaur et al. [1998] *supra*).

For substrate competition assays, end-labeled *CX/AY-31mer was generated by annealing 3' end-labeled CX with unlabeled strand AY. Unlabeled non-specific (non-mismatch) competitor GX/CY-31mer was made by annealing strand GX to strand CY resulting in a duplex oligonucleotide with a C/A base pair instead of a G/G mispair. CPD-30mer, a well-characterized substrate for Uve1p, was employed as an unlabeled, specific competitor. 3' end-labeled *CX/AY-31mer (0.1 pmol) was incubated with 100 ng of purified GΔ228-Uve1p and increasing amounts (0.1-2.0 pmol) of either specific (CPD-30mer) or non-specific (GX/CY-31mer) competitor. The competition reactions were processed and analyzed on 20% denaturing gels as described above. The DNA species corresponding to the uncleaved *GX/GY-31mer and the DNA strand scission products were quantified by phosphorimager analysis (Molecular Dynamics model 445SI).

Example 11. Mutation Frequencies Assayed by Canavanine Resistance.

To determine sensitivity to L-canavanine, 10 ml of PMALU^g was inoculated with 100 μl of the indicated saturated culture and grown to mid-log phase at 25°C. 200 cells were plated onto PMALU^g plates with varying concentrations of L-canavanine sulfate (0, 0.075, 0.22, 0.75, 2.2, 7.5, 22, and 75 μg/ml) and incubated at 30°C. Colonies were counted after four days and viability was normalized against the 0 g/ml plate for each strain. Colony formation assays were conducted for each strain by plating 10^7 cells from saturated cultures onto PMALU^g plates supplemented with 75 μg/ml L-canavanine sulfate. Colonies were

counted after eight days incubation at 30°C. Mean mutation frequencies were calculated using the method of the median as described by Lea and Coulson (1943) *J. Genet.* **49**:264-284.

Table 1A. Damaged Oligonucleotide Substrates Used in This Study.

cis-syn cyclobutane pyrimidine dimers (cs-CPDs), *trans-syn* I CPD (tsI-CPD), *trans-syn* II CPD (tsII-CPD), (6-4) photoproducts (6-4PP), a Dewar isomer (Dewar), a platinum DNA diadduct (Pt-GG), uracil (U), dihydrouracil (DHU), abasic site (AP), inosine (I), xanthine (Xn) and 8-oxoguanine (8-oxoG).

Substrate	Damaged oligonucleotide sequence 5' to 3'	Adduct	[a]Opposite base(s).	SEQ ID NO.:
A: cs-CPD-30mer	CATGCCTGCACGAAT^TAAGCAATTCGTAAT	cs-CPD	AA	13
B: UD-30mer	CATGCCTGCACGAATTAAGCAATTCGTAAT	undamaged	AA	14
C: cs-CPD-49mer	AGCTACCATGCCTGCACGAAT^TAAGCAATTCGTAATCATGGTCATAGCT	cs-CPD	AA	15
D: tsI-CPD-49mer	"	tsI-CPD	AA	16
E: tsII-CPD-49mer	"	tsII-CPD	AA	17
F: 6-4PP-49mer	"	6-4PP	AA	18
G: Dewar-49mer	"	Dewar	AA	19
H: Pt-GG-32mer	TCCCTCCTTCCTTCCG*G*CCCTCCTTCCCCTTC	Pt-GG	CC	20
I: U-37mer	CTTGGACTGGATGTCGGCACXAGCGGATACAGGAGCA	U	A/G	21
J: DHU-37mer	"	DHU	A/G	22
K: AP-37mer	"	AP	A/G	23
L: I-31mer	TGCAGGTCGACTXAGGAGGATCCCCGGGTAC	I	T/C	24
M: Xn-31mer	"	Xn	T/C	25
N: 8-oxoG-37mer	CTTGGACTGGATGTCGGCACXAGCGGATACAGGAGCA	8-oxoG	A/T/G/C	26

[a]denotes the bases that are placed opposite to the lesions on the complementary DNA strand.

^, *, X represent a UV induced dimer between two adjacent thymines, a cisplatin induced diadduct between two adjacent guanines and position at which the adducts U, DHU, AP, I, Xn and 8-oxoG are incorporated into the oligonucleotide substrates, respectively.

Table 1B: Base Mismatch and CPD-Containing Oligonucleotides Used in This Study.

Oligo Name	Sequence	Strand Designation	SEQ ID NO:
XY-31mer[1]	5' GTACCCGGGGATCCTCCXAGTCGACCTGCA 3' 3' CATGGGCCCCTAGGAGGYTCAGCTGGACGT 5'	GX, AX, TX, CX: X=G, A, T, C GY, AY, TY, CY: Y=G, A, T, C	27
CX/AY-41mer	5' CGTTAGCATGCCTGCACGAACTAAGCAATTCGTAATGCATT 3' 3' GCAATCGTACGGACGTGCTTAATTCGTTAAGCATTACGTAA 5'	CX AY	28
C(6)/A-41mer[2]	5' CGTTACAAGTCCGTCACGAATTAAGCAATTCGTAACGCATT 3' 3' GCAATGTTCAGGCAGTGCTTAATTCGTTAAGCATTACGTAA 5'	C(6) A(36)	29
C(11)/A-41mer	5' CGTTACAAGTCCGTCACGAATTAAGCAATTCGTAACGCATT 3' 3' GCAATGTTCAGGCAGTGCTTAATTCGTTAAACATTGCGTAA 5'	C(11) A(31)	30
C(16)/A-41mer	5' CGTTACAAGTCCGTCACGAATTAAGCAATTCGTAACGCATT 3' 3' GCAATGTTCAGGCAGTGCTTAATTCATTAAGCATTGCGTAA 5'	C(16) A(26)	31
C(22)/A-41mer	5' CGTTACAAGTCCGTCACGACTTAAGCAATTCGTAACGCATT 3' 3' GCAATGTTCAGGCAGTGCTAAATTCGTTAAGCATTGCGTAA 5'	C(22) A(20)	32
C(27)/A-41mer	5' CGTTACAAGTCCGTCACGAATTAAGCAATTCGTAACGCATT 3' 3' GCAATGTTCAGGCAATGCTTAATTCGTTAAGCATTGCGTAA 5'	C(27) A(15)	33
C(32)/A-41mer	5' CGTTACAAGCCCGTCACGAATTAAGCAATTCGTAACGCATT 3' 3' GCAATGTTCAGGCAGTGCTTAATTCGTTAAGCATTGCGTAA 5'	C(32) A(10)	34
C(37)/A-41mer	5' CGTTCCAAGTCCGTCACGAATTAAGCAATTCGTAACGCATT 3' 3' GCAAAGTTCAGGCAGTGCTTAATTCGTTAAGCATTGCGTAA 5'	C(37) A(5)	35
CPD-30mer[3]	5' CATGCCTGCACGAAT^TAAGCAATTCGTAAT 3' 3' GTACGGACGTGCTTA ATTCGTTAAGCATTA 5'	30 D 30 C	13

[1]Series of 16 different duplex oligos containing all possible base pair/mispair combinations between G, A, T, and C.
In text, * denotes labeled strand (e.g. *CX/AY-31mer corresponds to C/A mismatch with the C-containing X strand as the labeled strand).
[2]C/A mismatch oligos designated by base position of the mismatched C from the 3' terminus.
[3]CPD-30mer contains a cyclobutane pyrimidine dimer designated as T^T.

Table 2. Activity of Uve1p on Oligonucleotide Substrates Containing Uracil, Dihydrouracil and AP sites

Protein	U/G	U/A	DHU/G	DHU/A	AP/G	AP/A
[a]Positive control	90-100	50-60	70-80	15-20	90-100	90-100
GΔ228-Uve1p	8-12	1-5	37-42	10-15	90-100	90-100
GST	1-5	1-5	1-5	1-5	1-5	1-5

The percent of substrate converted into total DNA cleavage products formed when the DNA damage lesion is base paired with a G or an A in the complementary strand. Details of experiments are outlined in Example 10.

[a]Positive control: when analyzing U 37mer, uracil DNA glycosylase (UDG) was used as a positive control; for assays involving DHU 37mer, the *S. cerevisiae* endonuclease III-like homolog Ntg1 was used as a positive control; *E. coli* endonuclease IV was used as a positive control for AP endonuclease activity.

Table 3. Uve1p Cleavage Efficiency on Different Substrates.

Substrate	Percent Cleavage[a]
cs-CPD 49mer	89
tsI-CPD 49mer	75
tsII-CPD 49mer	75
6-4PP 49mer	71
Dewar	83
AP 37mer	12.5
DHU 37mer	3
Pt-GG 32mer	2.5
U 37mer	1
8-oxoG 37mer	0
I 31mer	0
Xn 31mer	0

[a]The percent cleavage was calculated by quantifying the amount of Uve1p-mediated cleavage product formed when 300 ng of affinity-purified GΔ228-Uve1p was incubated with ~150 fmol of each substrate.

Table 4. Spontaneous Mutation Rates of *uve1* and *pms1* Null Mutants

Genotype	Distribution of canavanine-resistant colonies/plate				Median no. of colonies/10^7 cells	Calculated mutation frequency (mean ± SE)
	0-2	3-34	35-86	>86		
Wild type	18	16	2	0	2.5	$1.5 \times 10^{-7} \pm 2.5 \times 10^{-8}$
uve1::ura4$^+$	4	14	8	10	34.5	$9.7 \times 10^{-7} \pm 4.2 \times 10^{-8}$
pms1::ura4$^+$	0	8	10	18	86.5	$2.0 \times 10^{-6} \pm 5.0 \times 10^{-8}$

Table 5. Nucleotide Sequence Encoding GST-Full-length UVDE (SEQ ID NO:1)

```
   1  atgaccaagt tacctatact aggttattgg aaaaattaag ggccttgtgc
  51  aacccactcg acttcttttg gaatatcttg aagaaaaata tgaagagcat
 101  ttgtatgagc gcgatgaagg tgataaatgg cgaaacaaaa agtttgaatt
 151  gggtttggag tttcccaatc ttccttatta tattgatggt gatgttaaat
 201  taacacagtc tatggccatc atacgttata tagctgacaa gcacaacatg
 251  ttggttggtt gtccaaaaga gcgtgcagag atttcaatgc ttgaaggagc
 301  ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact
 351  ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa
 401  atgttcgaag atcgtttatg tcataaaaca tatttaaatg ttgaccatgt
 451  aacccatcct gacttcatgt tgtatgacgc tcttgatgtt gttttataca
 501  tggacccaat gtgcctggat gcgttcccaa aattagtttg ttttaaaaaa
 551  cgtattgaag ctatcccaca aattgataag tacttgaaat ccagcaagta
 601  tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc
 651  atcctccaaa atcggatcat gtggttccgc gtggatccat gcttaggcta
 701  ttgaaacgaa atattcaaat ctctaaacgc attgttttca ccatattaaa
 751  acaaaaggca tttaaaggta atcatccttg tgtaccgtcg gtttgtacca
 801  ttacttactc tcgttttcat tgtttacccg atacccttaa aagtttactt
 851  ccaatgagct caaaaaccac actctcaatg ttaccgcaag ttaatatcgg
 901  tgcgaattca ttctctgccg aaacaccagt cgacttaaaa aaagaaaatg
 951  agactgagtt agctaatatc agtggacctc acaaaaaaag tacttctacg
1001  tctacacgaa agagggcacg tagcagtaaa aagaaagcga cagattctgt
1051  ttccgataaa attgatgagt ctgttgcgtc ctatgattct tcaactcatc
1101  ttaggcgatc gtcgagatca aaaaaaccgg tcaactacaa ttcctcgtca
```

Table 5. Continued

1151 gaatccgaat cggaggagca aattagtaaa gctactaaaa aagttaaaca
1201 aaaagaggaa gaggagtatg ttgaagaagt cgacgaaaag tctcttaaaa
1251 atgaaagtag ctctgacgag ttcgaaccgg ttgtgccgga acagttggaa
1301 actccaattt ctaaacgaag acggtctcgt tcttctgcaa aaaatttaga
1351 aaaagaatct acaatgaatc ttgatgatca tgctccacga gagatgtttg
1401 attgtttgga caaacccata ccctggcgag gacgattggg gtatgcttgt
1451 ttgaatacta ttttaaggtc aatgaaggag agggtttttt gttcacgcac
1501 ctgccgaatt acaaccattc aacgtgatgg gctcgaaagt gtcaagcagc
1551 taggtacgca aaatgtttta gatttaatca aattggttga gtggaatcac
1601 aactttggca ttcacttcat gagagtgagt tctgatttat ttcctttcgc
1651 aagccatgca aagtatggat ataccccttga atttgcacaa tctcatctcg
1701 aggaggtggg caagctggca aataaatata atcatcgatt gactatgcat
1751 cctggtcagt acacccagat agcctctcca cgagaagtcg tagttgattc
1801 ggcaatacgt gatttggctt atcatgatga aattctcagt cgtatgaagt
1851 tgaatgaaca attaaataaa gacgctgttt taattattca ccttggtggt
1901 acctttgaag gaaaaaaaga aacattggat aggtttcgta aaaattatca
1951 acgcttgtct gattcggtta aagctcgttt agttttagaa aacgatgatg
2001 tttcttggtc agttcaagat ttattacctt tatgccaaga acttaatatt
2051 cctctagttt tggattggca tcatcacaac atagtgccag gaacgcttcg
2101 tgaaggaagt ttagatttaa tgccattaat cccaactatt cgagaaacct
2151 ggacaagaaa gggaattaca cagaagcaac attactcaga atcggctgat
2201 ccaacggcga tttctgggat gaaacgacgt gctcactctg atagggtgtt
2251 tgactttcca ccgtgtgatc ctacaatgga tctaatgata gaagctaagg
2301 aaaaggaaca ggctgtattt gaattgtgta gacgttatga gttacaaaat
2351 ccaccatgtc ctcttgaaat tatggggcct gaatacgatc aaactcgaga
2401 tggatattat ccgcccggag ctgaaaagcg tttaactgca agaaaaaggc
2451 gtagtagaaa agaagaagta gaagaggatg aaaaataaaa at

Table 6. Deduced Amino Acid Sequences of GST-Full-length UVDE (SEQ ID NO:2)

```
  1  mtklpilgyw kikglvqptr llleyleeky eehlyerdeg dkwrnkkfel
 51  glefpnlpyy idgdvkltqs maiiryiadk hnmlggcpke raeismlega
101  vldirygvsr iayskdfetl kvdflsklpe mlkmfedrlc hktylngdhv
151  thpdfmlyda ldvvlymdpm cldafpklvc fkkrieaipq idkylkssky
201  iawplqgwqa tfgggdhppk sdhlvprgsm lrllkrniqi skrivftilk
251  qkafkgnhpc vpsvctitys rfhclpdtlk sllpmssktt lsmlpqvnig
301  ansfsaetpv dlkkenetel anisgphkks tststrkrar sskkkatdsv
351  sdkidesvas ydssthlrrs srskkpvnyn ssseseseeq iskatkkvkq
401  keeeeyveev dekslkness sdefepvvpe qletpiskrr rsrssaknle
451  kestmnlddh apremfdcld kpipwrgrlg yaclntilrs mkervfcsrt
501  crittiqrdg lesvkqlgtq nvldliklve wnhnfgihfm rvssdlfpfa
551  shakygytle faqshleevg klankynhrl tmhpgqytqi asprevvvds
601  airdlayhde ilsrmklneq lnkdavliih lggtfegkke tldrfrknyq
651  rlsdsvkarl vlenddvsws vqdllplcqe lniplvldwh hhnivpgtlr
701  egsldlmpli ptiretwtrk gitqkqhyse sadptaisgm krrahsdrvf
751  dfppcdptmd lmieakekeq avfelcrrye lqnppcplei mgpeydqtrd
801  gyyppgaekr ltarkrrsrk eeveedek
```

Table 7. Nucleotide Sequence Encoding Δ228-UVDE (SEQ ID NO:3)

```
   1  gatgatcatg ctccacgaga gatgtttgat tgtttggaca aacccatacc
  51  ctggcgagga cgattggggt atgcttgttt gaatactatt ttaaggtcaa
 101  tgaaggagag ggtttttttgt tcacgcacct gccgaattac aaccattcaa
 151  cgtgatgggc tcgaaagtgt caagcagcta ggtacgcaaa atgttttaga
 201  tttaatcaaa ttggttgagt ggaatcacaa ctttggcatt cacttcatga
 251  gagtgagttc tgatttattt cctttcgcaa gccatgcaaa gtatggatat
 301  acccttgaat ttgcacaatc tcatctcgag gaggtgggca agctggcaaa
 351  taaatataat catcgattga ctatgcatcc tggtcagtac acccagatag
 401  cctctccacg agaagtcgta gttgattcgg caatacgtga tttggcttat
 451  catgatgaaa ttctcagtcg tatgaagttg aatgaacaat taaataaaga
 501  cgctgtttta attattcacc ttggtggtac ctttgaagga aaaaaagaaa
 551  cattggatag gtttcgtaaa aattatcaac gcttgtctga ttcggttaaa
 601  gctcgtttag ttttagaaaa cgatgatgtt tcttggtcag ttcaagattt
 651  attaccttta tgccaagaac ttaatattcc tctagttttg gattggcatc
 701  atcacaacat agtgccagga acgcttcgtg aaggaagttt agatttaatg
 751  ccattaatcc caactattcg agaaacctgg acaagaaagg gaattacaca
 801  gaagcaacat tactcagaat cggctgatcc aacggcgatt tctgggatga
 851  aacgacgtgc tcactctgat agggtgtttg actttccacc gtgtgatcct
 901  acaatggatc taatgataga agctaaggaa aaggaacagg ctgtatttga
 951  attgtgtaga cgttatgagt tacaaaatcc accatgtcct cttgaaatta
1001  tggggcctga atacgatcaa actcgagatg gatattatcc gcccggagct
1051  gaaaagcgtt taactgcaag aaaaaggcgt agtagaaaag aagaagtaga
1101  agaggatgaa aaataaaaat ccgtcatact ttttgattta tggcataatt
1151  tagccatctc c
```

Table 8. Deduced Amino Acid Sequence of Δ228-UVDE (SEQ ID NO:4)

```
  1  ddhapremfd cldkpipwrg rlgyaclnti lrsmkervfc srtcrittiq
 51  rdglesvkql gtqnvldlik lvewnhnfgi hfmrvssdlf pfashakygy
101  tlefaqshle evgklankyn hrltmhpgqy tqiasprevv vdsairdlay
151  hdeilsrmkl neqlnkdavl iihlggtfeg kketldrfrk nyqrlsdsvk
201  arlvlenddv swsvqdllpl cqelniplvl dwhhhnivpg tlregsldlm
251  pliptiretw trkgitqkqh ysesadptai sgmkrrahsd rvfdfppcdp
301  tmdlmieake keqavfelcr ryelqnppcp leimgpeydq trdgyyppga
351  ekrltarkrr srkeeveede k
```

Table 9. Nucleotide Sequence Encoding GST-Δ228-UVDE (SEQ ID NO:5)

```
   1  atgaccaagt tacctatact aggttattgg aaaaattaag ggccttgtgc
  51  aacccactcg acttcttttg gaatatcttg aagaaaaata tgaagagcat
 101  ttgtatgagc gcgatgaagg tgataaatgg cgaaacaaaa agtttgaatt
 151  gggtttggag tttcccaatc ttccttatta tattgatggt gatgttaaat
 201  taacacagtc tatggccatc atacgttata tagctgacaa gcacaacatg
 251  ttggttggtt gtccaaaaga gcgtgcagag atttcaatgc ttgaaggagc
 301  ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact
 351  ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa
 401  atgttcgaag atcgtttatg tcataaaaca tatttaaatg ttgaccatgt
 451  aacccatcct gacttcatgt tgtatgacgc tcttgatgtt gttttataca
 501  tggacccaat gtgcctggag gcgttcccaa aattagtttg ttttaaaaaa
 551  cgtattgaag ctatcccaca aattgataag tacttgaaat ccagcaagta
 601  tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc
 651  atcctccaaa atcggatcat ctggttccgc gtggatccga tgatcatgct
 701  ccacgagaga tgtttgattg tttggacaaa cccataccct ggcgaggacg
 751  attggggtat gcttgtttga atactatttt aaggtcaatg aaggagaggg
 801  ttttttgttc acgcacctgc cgaattacaa ccattcaacg tgatgggctc
 851  gaaagtgtca agcagctagg tacgcaaaat gttttagatt taatcaaatt
 901  ggttgagtgg aatcacaact ttggcattca cttcatgaga gtgagttctg
 951  atttatttcc tttcgcaagc catgcaaagt atggatatac ccttgaattt
1001  gcacaatctc atctcgagga ggtgggcaag ctggcaaata aatataatca
1051  tcgattgact atgcatcctg gtcagtacac ccagatagcc tctccacgag
1101  aagtcgtagt tgattcggca atacgtgatt tggcttatca tgatgaaatt
```

Table 9. Continued

```
1151  ctcagtcgta tgaagttgaa tgaacaatta aataaagacg ctgttttaat
1201  tattcacctt ggtggtacct ttgaaggaaa aaaagaaaca ttggataggt
1251  ttcgtaaaaa ttatcaacgc ttgtctgatt cggttaaagc tcgtttagtt
1301  ttagaaaacg atgatgtttc ttggtcagtt caagatttat tacctttatg
1351  ccaagaactt aatattcctc tagttttgga ttggcatcat cacaacatag
1401  tgccaggaac gcttcgtgaa ggaagtttag atttaatgcc attaatccca
1451  actattcgag aaacctggac aagaaaggga attacacaga agcaacatta
1501  ctcagaatcg gctgatccaa cggcgatttc tgggatgaaa cgacgtgctc
1551  actctgatag ggtgtttgac tttccaccgt gtgatcctac aatggatcta
1601  atgatagaag ctaaggaaaa ggaacaggct gtatttgaat tgtgtagacg
1651  ttatgagtta caaaatccac catgtcctct tgaaattatg gggcctgaat
1701  acgatcaaac tcgagatgga tattatccgc ccggagctga aaagcgttta
1751  actgcaagaa aaaggcgtag tagaaaagaa gaagtagaag aggatgaaaa
1801  ataaggatcc c
```

Table 10. Deduced Amino Acid Sequence of GST-Δ228-UVDE (SEQ ID NO:6)

```
  1  mtklpilgyw kikglvqptr llleyleeky eehlyerdeg dkwrnkkfel
 51  glefpnlpyy idgdvkltqs maiiryiadk hnmlggcpke raeismlega
101  vldirygvsr iayskdfetl kvdflsklpe mlkmfedrlc hktylngdhv
151  thpdfmlyda ldvvlymdpm cldafpklvc fkkrieaipq idkylkssky
201  iawplqgwqa tfgggdhppk sdhlvprgsd dhapremfdc ldkpipwrgr
251  lgyaclntil rsmkervfcs rtcrittiqr dglesvkqlg tqnvldlikl
301  vewnhnfgih fmrvssdlfp fashakygyt lefaqshlee vgklankynh
351  rltmhpgqyt qiasprevvv dsairdlayh deilsrmkln eqlnkdavli
401  ihlggtfegk ketldrfrkn yqrlsdsvka rlvlenddvs wsvqdllplc
451  qelniplvld whhhnivpgt lregsldlmp liptiretwt rkgitqkqhy
501  sesadptais gmkrrahsdr vfdfppcdpt mdlmieakek eqavfelcrr
551  yelqnppcpl eimgpeydqt rdgyyppgae krltarkrrs rkeeveedek
```

Table 11. Nucleotide Sequence Encoding the GST Leader Sequence (SEQ ID NO:7)

```
   1  atgaccaagt tacctatact aggttattgg aaaaattaag ggccttgtgc
  51  aacccactcg acttcttttg gaatatcttg aagaaaaata tgaagagcat
 101  ttgtatgagc gcgatgaagg tgataaatgg cgaaacaaaa agtttgaatt
 151  gggtttggag tttcccaatc ttccttatta tattgatggt gatgttaaat
 201  taacacagtc tatggccatc atacgttata tagctgacaa gcacaacatg
 251  ttggttggtt gtccaaaaga gcgtgcagag atttcaatgc ttgaaggagc
 301  ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact
 351  ttgaaactct caaagttgat tttcttagca agctacctga aatgctgaaa
 401  atgttcgaag atcgtttatg tcataaaaca tatttaaatg ttgaccatgt
 451  aacccatcct gacttcatgt tgtatgacgc tcttgatgtt gttttataca
 501  tggacccaat gtgcctggat gcgttcccaa aattagtttg ttttaaaaaa
 551  cgtattgaag ctatcccaca aattgataag tacttgaaat ccagcaagta
 601  tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc
 651  atcctccaaa atcggatcat ctggttccgc gtggatcc
```

Table 12. Deduced Amino Acid Sequence of the GST Leader Polypeptide (SEQ ID NO:8)

```
  1  MTKLPILGYW KIKGLVQPTR LLLEYLEEKY EEHLYERDEG DKWRNKKFEL
 51  GLEFPNLPYY IDGDVKLTQS MAIIRYIADK HNMLGGCPKE RAEISMLEGA
101  VLDIRYGVSR IAYSKDFETL KVDFLSKLPE MLKMFEDRLC HKTYLNGDHV
151  THPDFMLYDA LDVVLYMDPM CLDAFPKLVC FKKRIEAIPQ IDKYLKSSKY
201  IAWPLQGWQA TFGGGDHPPK SDHLVPRGS
```

Table 13. Neospora crassa UVDE Homolog (Genbank Accession No. BAA 74539) (SEQ ID NO:36)

mpsrkskaaaldtpqsesstfsstldssapspamlrrsgrnilqpssekdrdhekrsgeelagrmmgkda
nghclregkeqeegvkmaieglarmerrlqratkrqkkqleedgipvpsvvsrfptapyhhkstnaeere
akepvlkthskdvereaeigvddvvkmepaatniiepedaqdaaergaarppavnssylplpwkgrlg
yaclntylrnskppifssrtcrmasivdhrhplqfedepehhlknkpdkskepqdelghkfvqelglanar
divkmlcwnekygirflrlssemfpfashpvhgyklapfasevlaeagrvaaelghrltthpgqftqlgsp
rkevvesairdleyhdellsllklpeqqnrdavmiihmggqfgdkaatlerfkrnyarlsqsckrrlvlend
dvgwtvhdllpvceelnipmvldyhhhnicfdpahlregtldisdpklqeriantwkrkgikqkmhyse
pcdgavtprhrrkhrprvmtlppcppdmdlmieakdkeqavfelmrtfklpgfekindmvpydrdde
nrpappvkapkkkkggkrkrttdeeaaepeevdtaaddvkdapegpkevpeeeramggpynrvyw
plgceewlkpkkrevkkgkvpeevedegefdg

Table 14. Bacillus subtilis UVDE Homolog (Genbank Accession No. Z 49782) (SEQ ID NO:37)

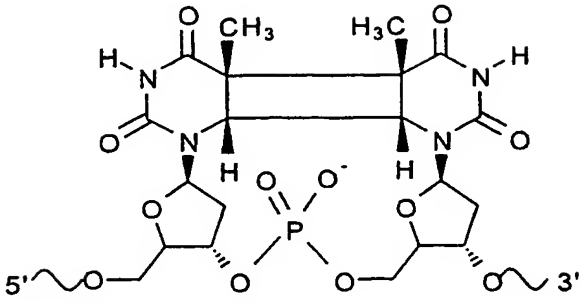
mifrfgfvsnamslwdaspaktltfarysklskterkealltvtkanlrntmrtlhyiighgiplyrfsssivpl
athpdvmwdfvtpfqkefreigelvkthqlrtsfhpnqftlftspkesvtknavtdmayhyrmleamgia
drsvinihiggaygnkdtataqfhqnikqlpqeikermtlenddktytteetlqvceqedvpfvfdfhhfy
anpddhadlnvalprmiktweriglqpkvhlsspkseqairshadyvdanfllpllerfrqwgtnidfmie
akqkdkallrlmdelssirgvkrigggalqwks

Table 15. Human UVDE Homolog (Genbank Accession No. AF 114784.1) (SEQ ID NO:38)

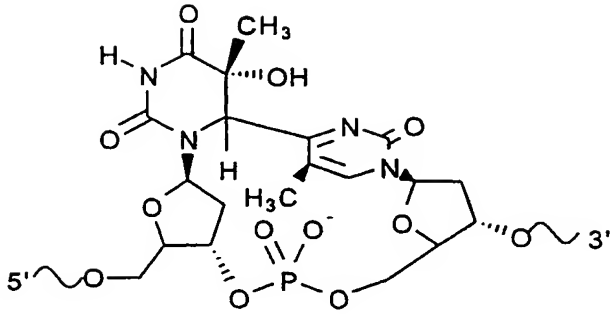
```
  1 mgttglesls lgdrgaaptv tsserlvpdp pndlrkedva melervgede eqmmikrsse
 61 cnpllqepia saqfgatagt ecrksvpcgw ervvkqrlfg ktagrfdvyf ispqglkfrs
121 ksslanylhk ngetslkped fdftvlskrg iksrykdcsm aaltshlqnq snnsnwnlrt
181 rskckkdvfm ppsssselqe srglsnftst hlllkedegv ddvnfrkvrk pkgkvtilkg
241 ipikktkkgc rkscsgfvqs dskresvcnk adaesepvaq ksqldrtvci sdagacgetl
301 svtseenslv kkkerslssg snfcseqkts giinkfcsak dsehnekyed tfleseeigt
361 kvevverkeh lhtdilkrgs emdnncsptr kdftgekifq edtiprtqie rrktslyfss
421 kynkealspp rrkafkkwtp prspfnivqe tlfhdpwkll iatiflnrts gkmaipvlwk
481 flekypsaev artadwrdvs ellkplglyd lraktivkfs deyltkqwky pielhgigky
541 gndsyrifcv newkqvhped hklnkyhdwl wenheklsls
```

Table 16. D. radiodurans UVDE Homolog (SEQ ID NO:39)

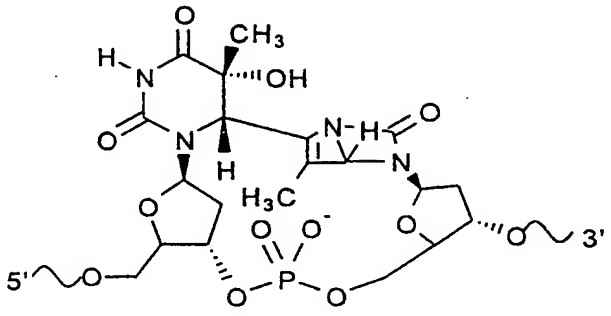
```
  1  QLGLVCLTVG PEVRFRTVTL SRYRALSPAE REAKLLDLYS SNIKTLRGAA

 51  DYCAAHDIRL YRLSSSLFPM LDLAGDDTGA AVLTHLAPQL LEAGHAFTDA

101  GVRLLMHPEQ FIVLNSDRPE VRESSVRAMS AHARVMDGLG LARTPWNLLL

151  LHGGKGGRGA ELAALIPDLP DPVRLRLGLE NDERAYSPAE LLPICEATGT

201  PLVFDAHHHV VHDKLPDQED PSVREWVLRA RATWQPPEWQ VVHLSNGIEG

251  PQDRRHSHLI ADFPSAYADV PWIEVEAKGK EEAIAALRLM APFK
```



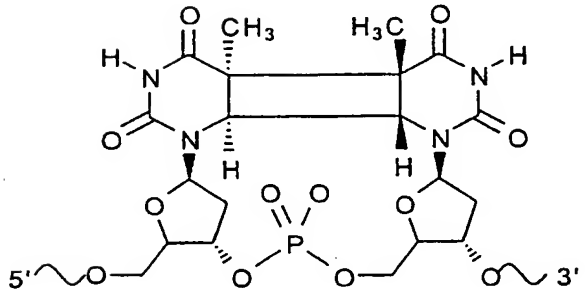
cs-CPD



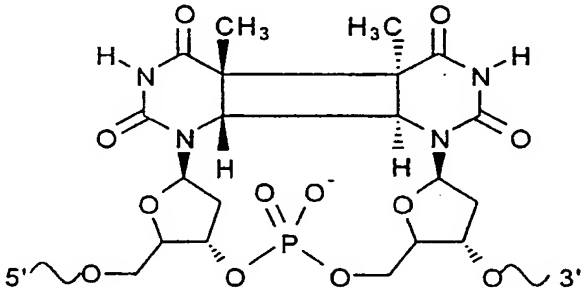
6-4PP



Dewar

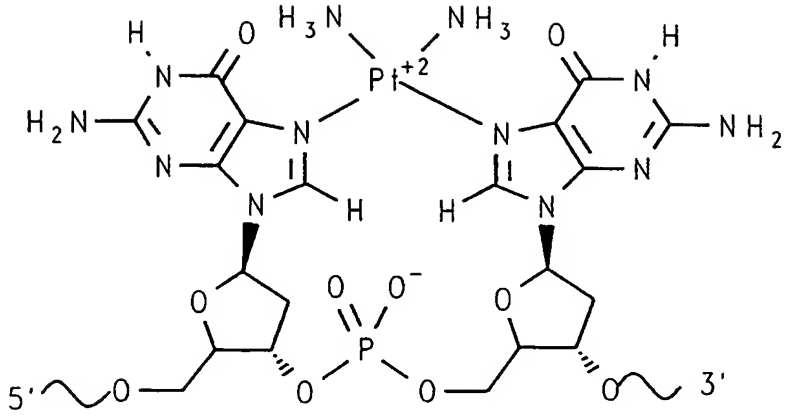


ts-I-CPD



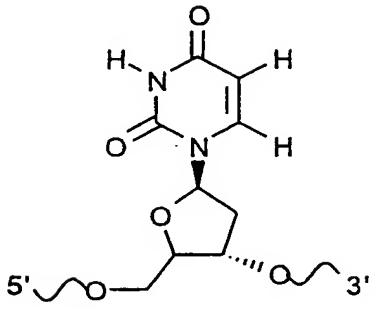
ts-II-CPD

SCHEME 1A

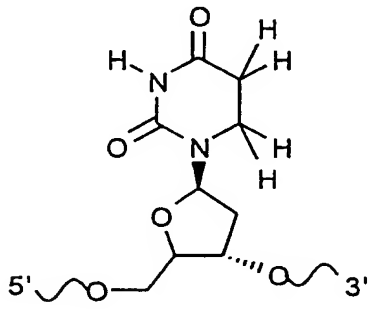


SCHEME 1B

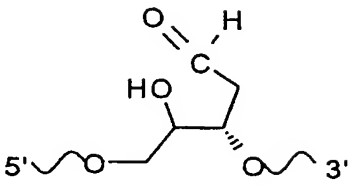
Pt-G G



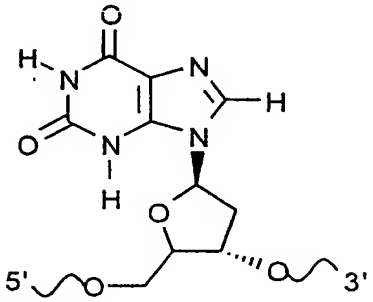
U



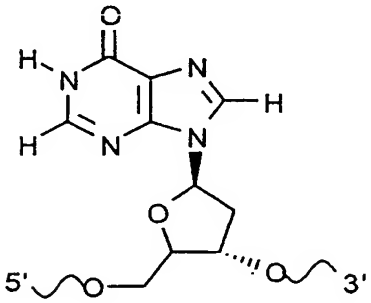
DHU



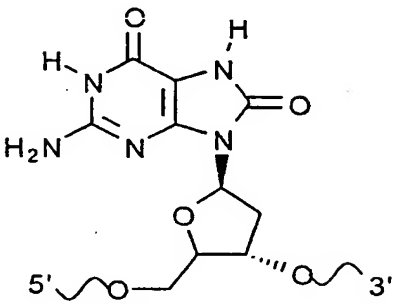
AP



Xn



I



8-oxoG

SCHEME 1C

WHAT IS CLAIMED IS:

1. A non-naturally occurring nucleic acid molecule comprising a portion which encodes a truncated ultraviolet damage endonuclease (Uve1p), said truncated Uve1p characterized by an amino acid sequence extending from a position between 329 and 479 as given in SEQ ID NO:2 and extending through amino acid 828 of SEQ ID NO:2.

2. The non-naturally occurring nucleic acid molecule of claim 1 encoding a stable truncated Uve1p characterized by an amino acid sequence as given in SEQ ID NO:2, amino acids 330 to 828.

3. The non-naturally occurring nucleic acid molecule of claim 1 encoding a stable truncated Uve1p characterized by an amino acid sequence as given in SEQ ID NO:2, amino acids 458 to 828.

4. The non-naturally occurring nucleic acid molecule of claim 1 encoding a stable truncated Uve1p characterized by an amino acid sequence as given in SEQ ID NO:2, amino acids 518 to 828.

5. The non-naturally occurring nucleic acid molecule of claim 3 encoding a stable truncated Uve1p, wherein said stable truncated Uve1p is encoded by a nucleotide sequence as given in SEQ NO:3.

6. The non-naturally occurring nucleic acid molecule of claim 1, wherein said nucleic acid molecule is a vector molecule.

7. A substantially purified stable truncated UV damage endonuclease (Uve1p) wherein said Uve1p has amino acid sequence as given in SEQ ID NO:2, wherein its amino-terminus is between about amino acid 329 and about amino acid 479, and extends through amino acid 828 of SEQ ID NO:2.

8. The substantially purified stable truncated Uvelp of claim 7 wherein its amino acid sequence is as given in SEQ ID NO:2, amino acid 458 through amino acid 828.

9. The substantially purified stable truncated Uvelp of claim 8 further comprising a polypeptide portion identified by an amino acid sequence as given in SEQ ID NO:8 covalently joined at its amino terminus.

10. The substantially purified stable truncated Uvelp of claim 7 wherein said Uvelp has an amino acid sequence as given in SEQ ID NO:2, amino acid 458 through amino acid 828.

11. The substantially purified stable truncated Uvelp of claim 10 further a polypeptide portion identified by an amino acid sequence as given in SEQ ID NO:8 covalently joined at its N-terminus.

12. A composition comprising a substantially purified stable truncated Uvelp of claim 7 and a pharmacologically acceptable carrier.

13. The composition of claim 12 wherein said truncated Uvelp has an amino acid sequence as given in SEQ ID NO:4.

14. The composition of claim 12 which is formulated for topical application to skin of a human or an animal.

15. The composition of claim 12 which is formulated for internal use in a human or an animal.

16. A method for cleavage of a double-stranded DNA molecule characterized by a distorted structure, wherein said distorted structure results from ultraviolet radiation damage, a photoproduct, an abasic site, mismatched nucleotide pairing, a platinum diadduct, an intercalated ,molecule, or alkylation of a nucleotide, said method

comprising the step of contacting a DNA molecule characterized by a distorted structure with a broadly specific DNA damage endonuclease selected from the group of endonucleases selected from the group consisting of an endonuclease identified by the amino acid sequence as given in SEQ ID NO:2, amino acids 230 to 828; a truncated stable truncated Uve1p identified by the amino acid sequence given in SEQ ID NO:4; the endonuclease identified by the amino acid sequence given in SEQ ID NO:36; the endonuclease identified by the amino acid sequence given in SEQ ID NO:37; the endonuclease identified by the amino acid sequence given in SEQ ID NO:38; the endonuclease identified by the amino acid sequence given in SEQ ID NO:39, under conditions allowing for enzymatic activity of said endonuclease.

17. The composition of claim 16 wherein said truncated Uve1p has an amino acid sequence as given in SEQ ID NO:4.

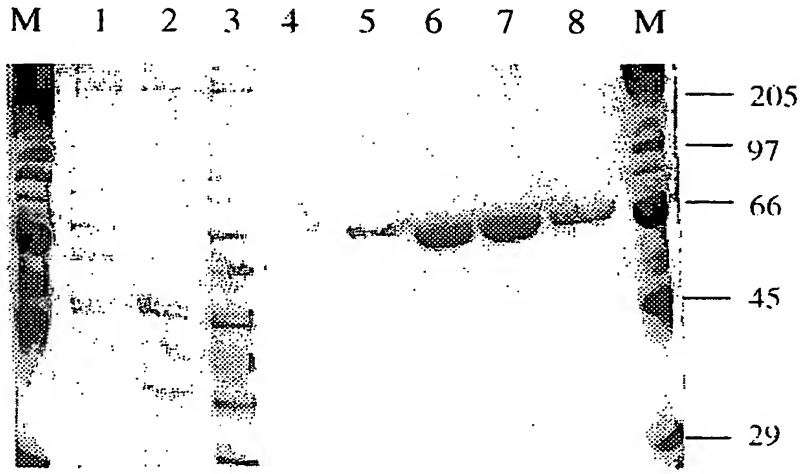


FIG. 1A

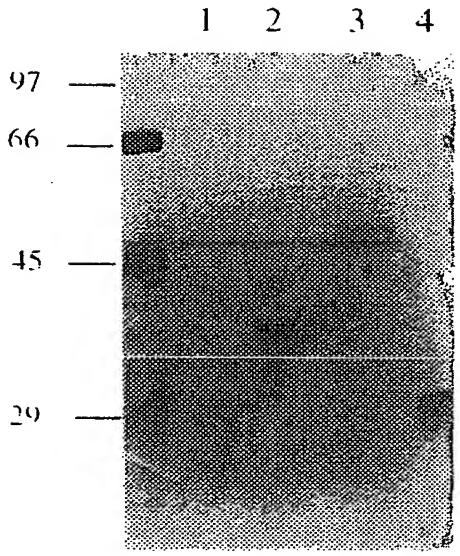


FIG. 1B

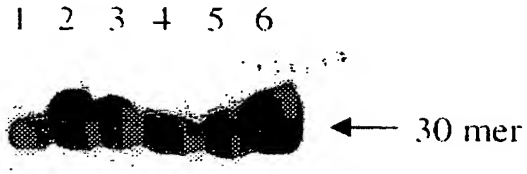


FIG. 1C

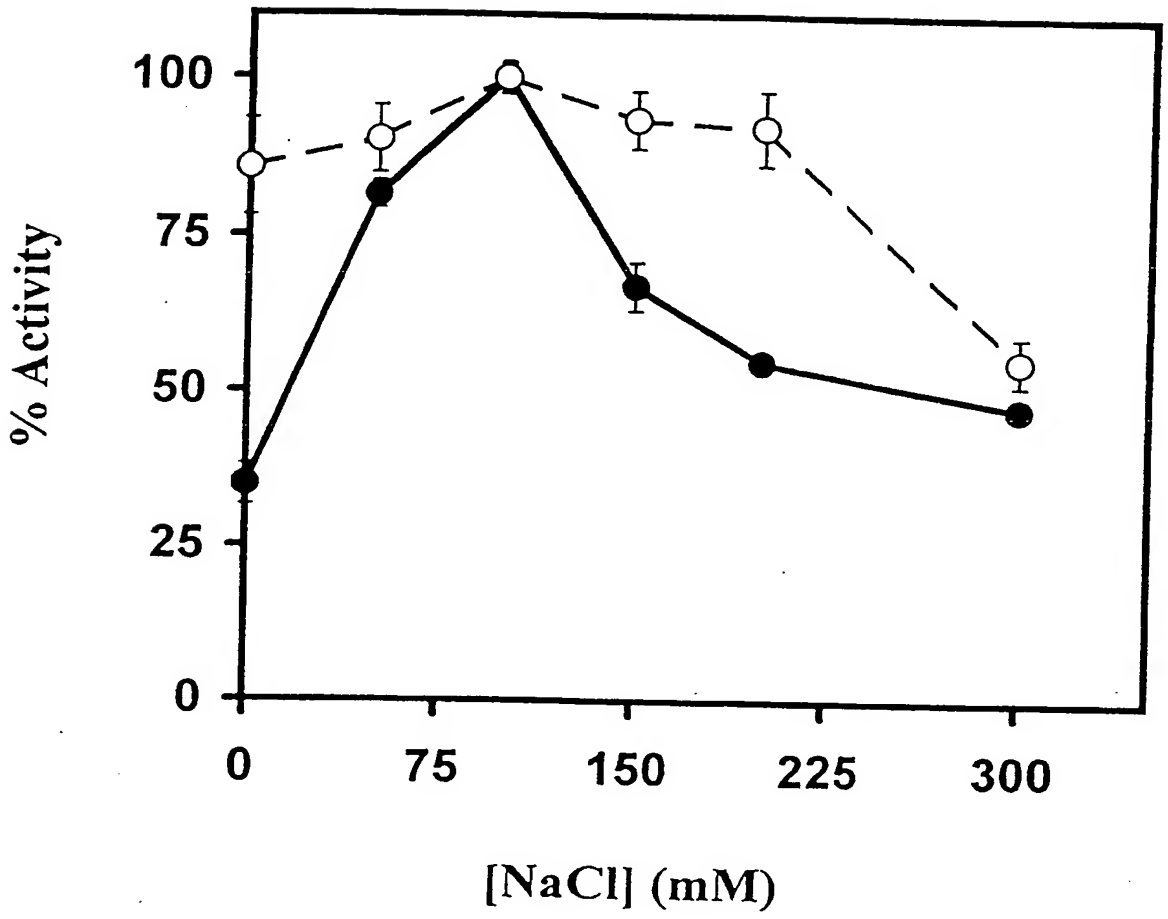


FIG. 2

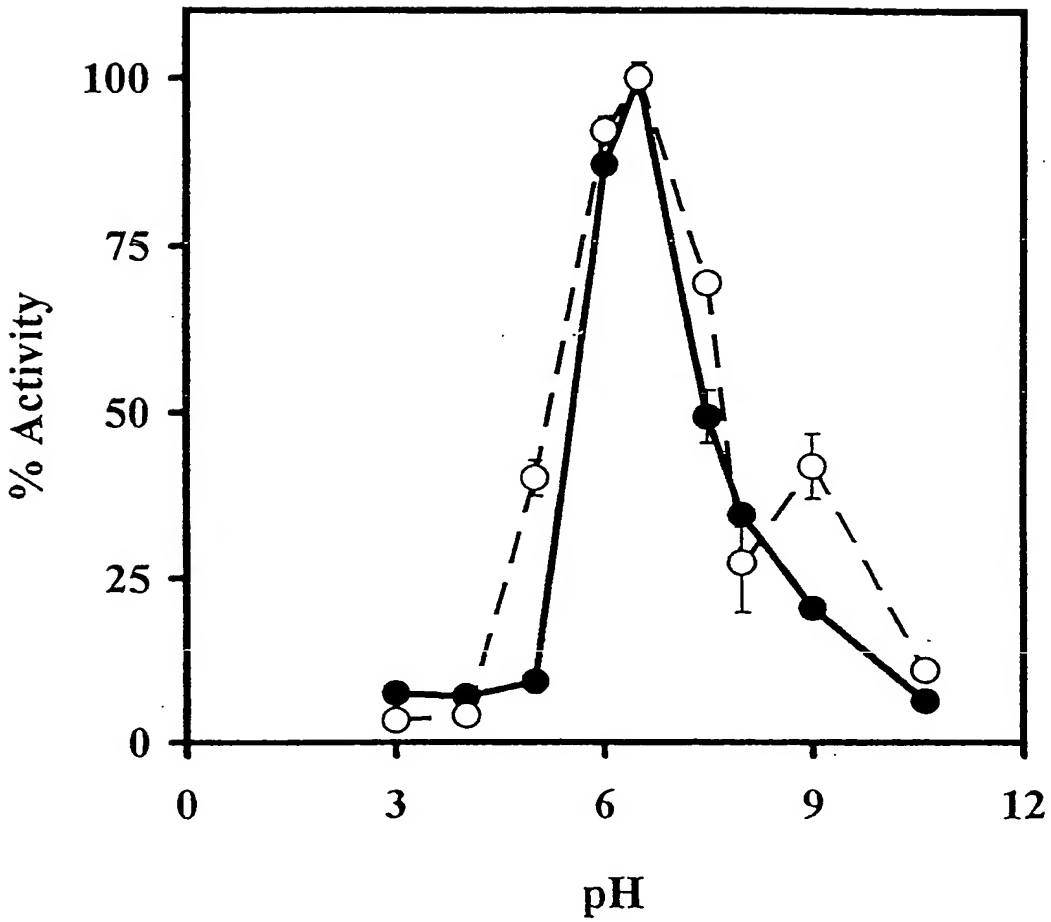


FIG. 3

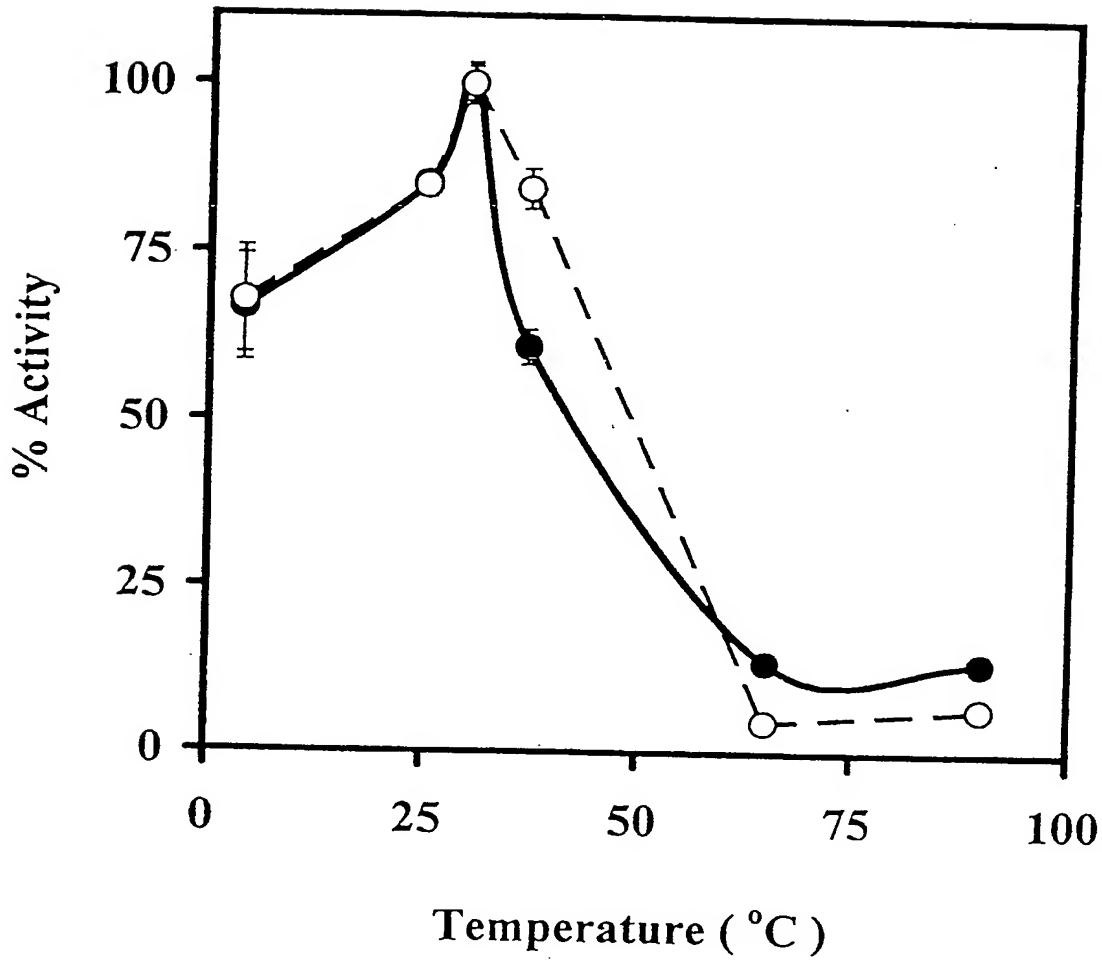


FIG. 4

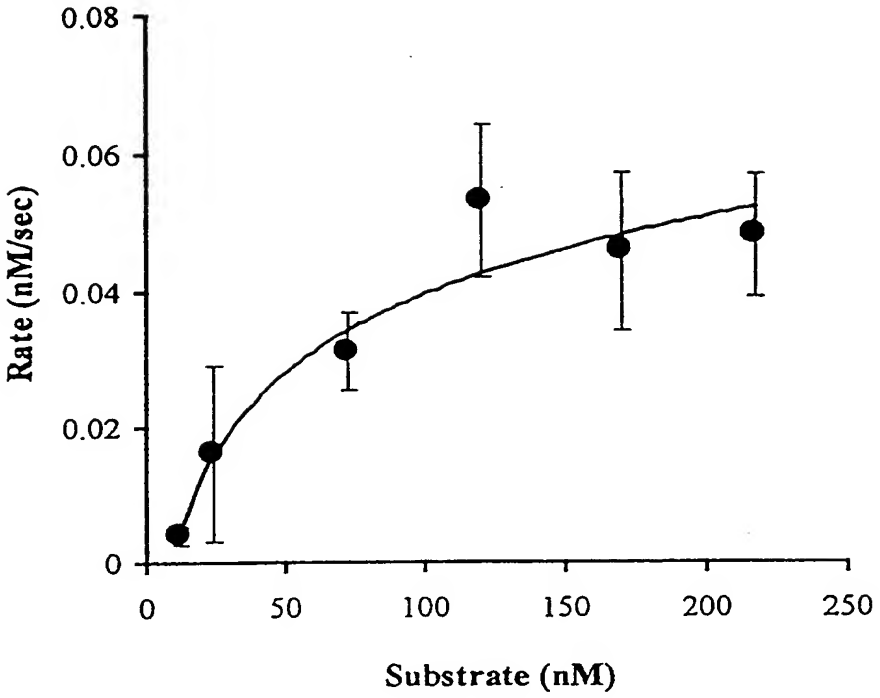


FIG. 5A

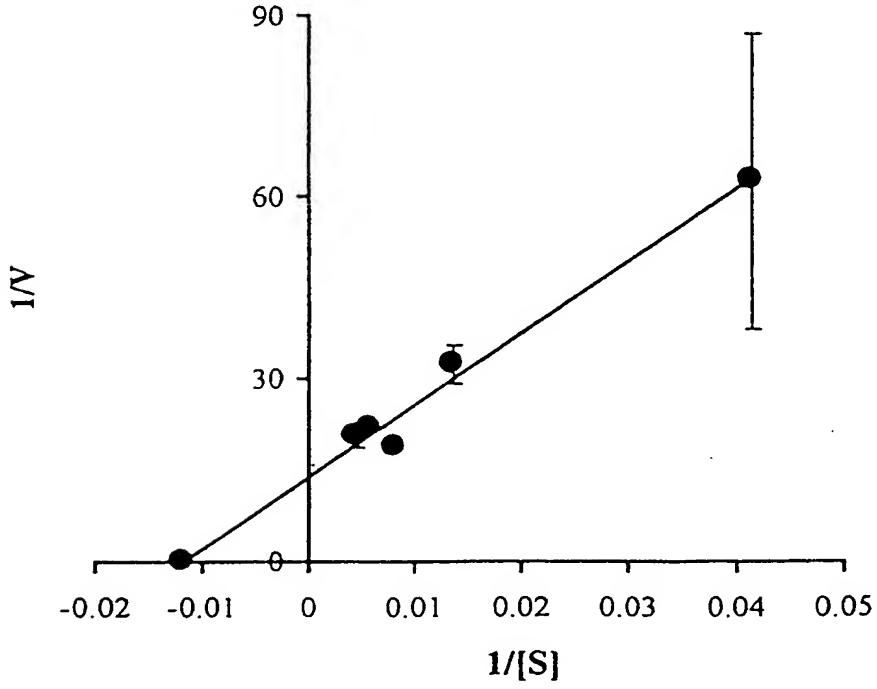


FIG. 5B

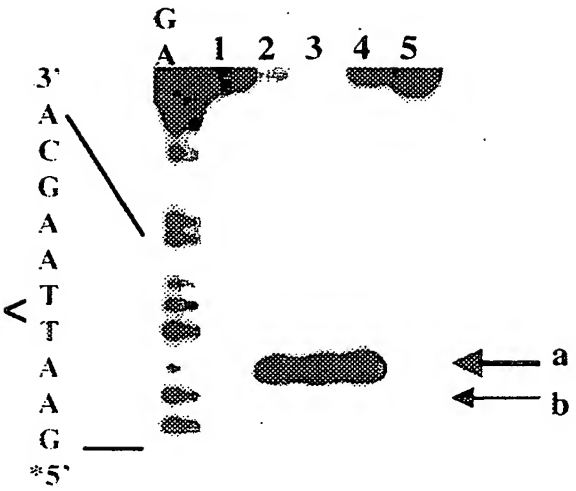


FIG. 6A

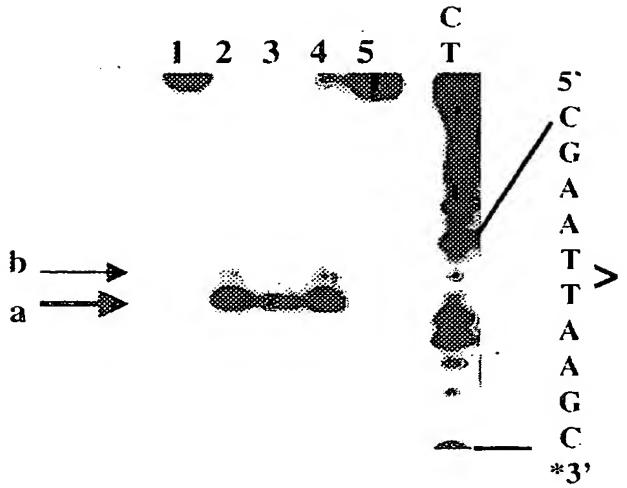


FIG. 6B

FIG. 7A

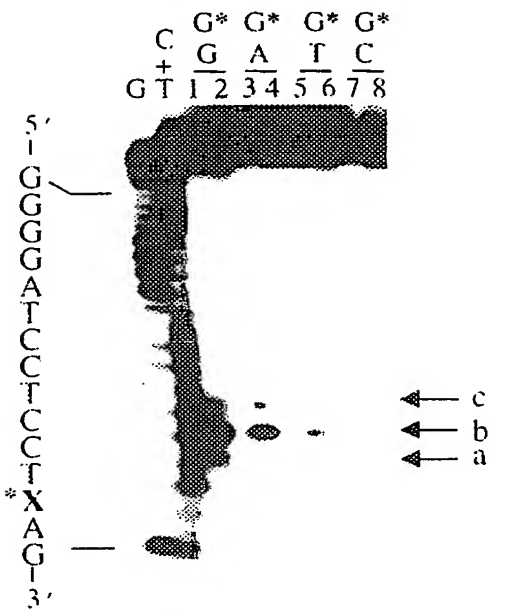


FIG. 7B

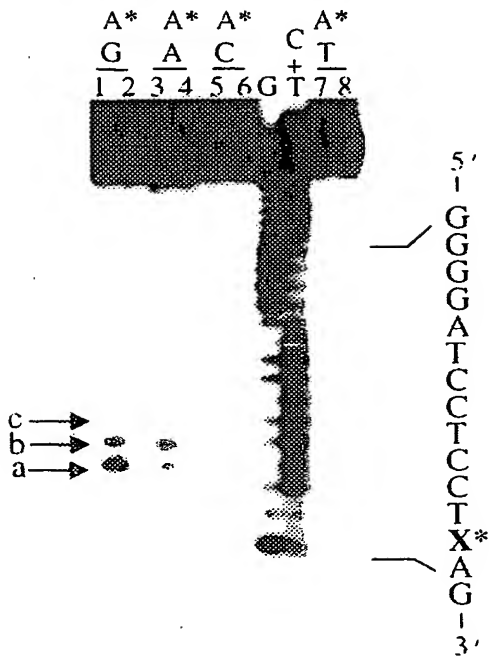


FIG. 7C

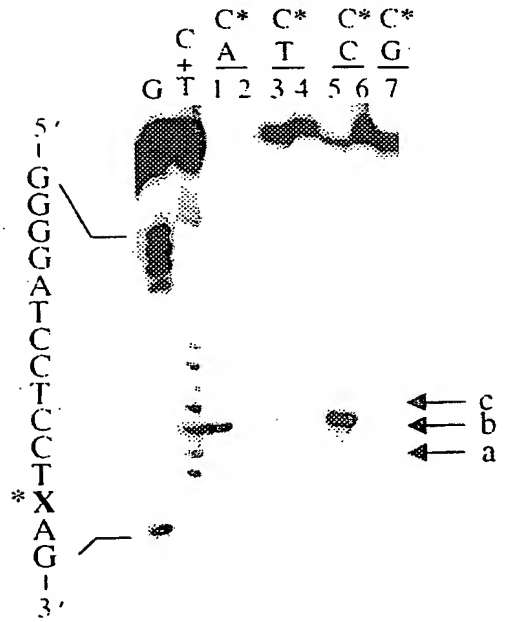
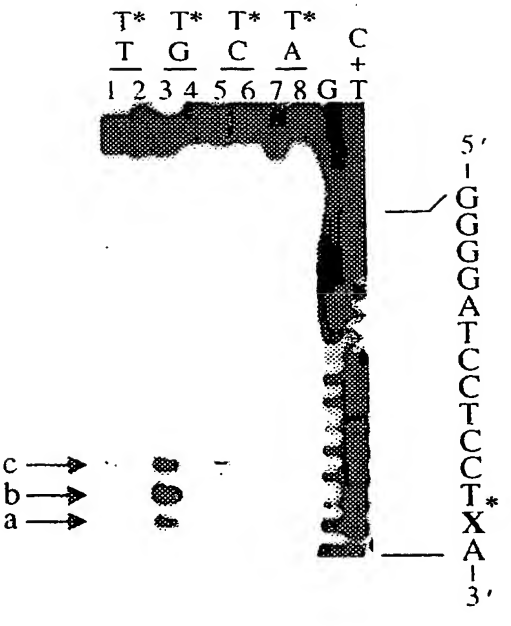


FIG. 7D



c b a

5′ GTACCCGGGGATCCTCCT**X**AGTCGACCTGCA* 3′
3′ CATGGGCCCCTAGGAGGA**Y**TCAGCTGGACGT 5′

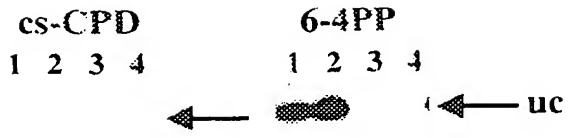


FIG. 8A

FIG. 8B

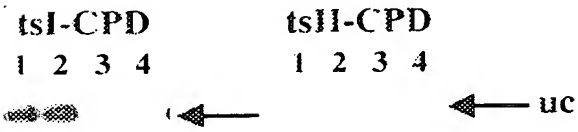
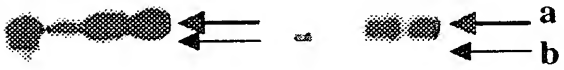


FIG. 8C

FIG. 8D



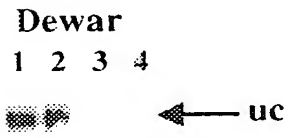
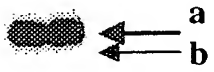
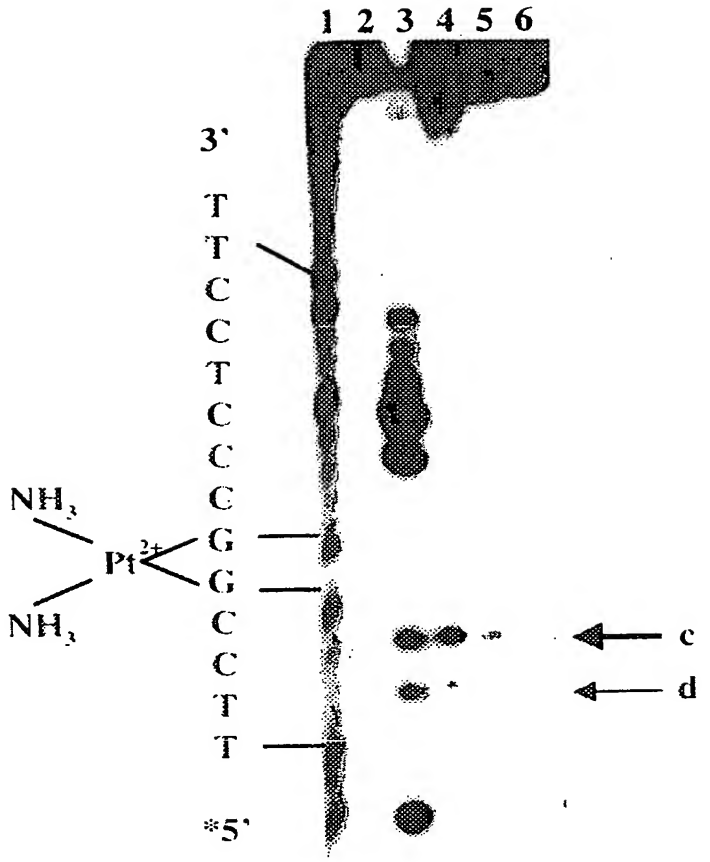


FIG. 8E





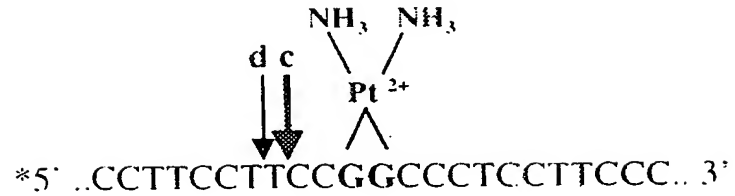


FIG. 9

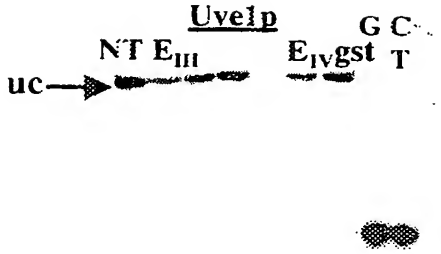
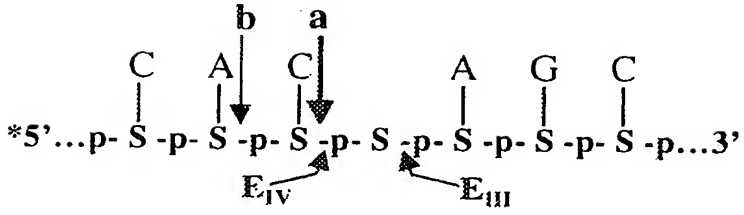


FIG. 10A

FIG. 10B



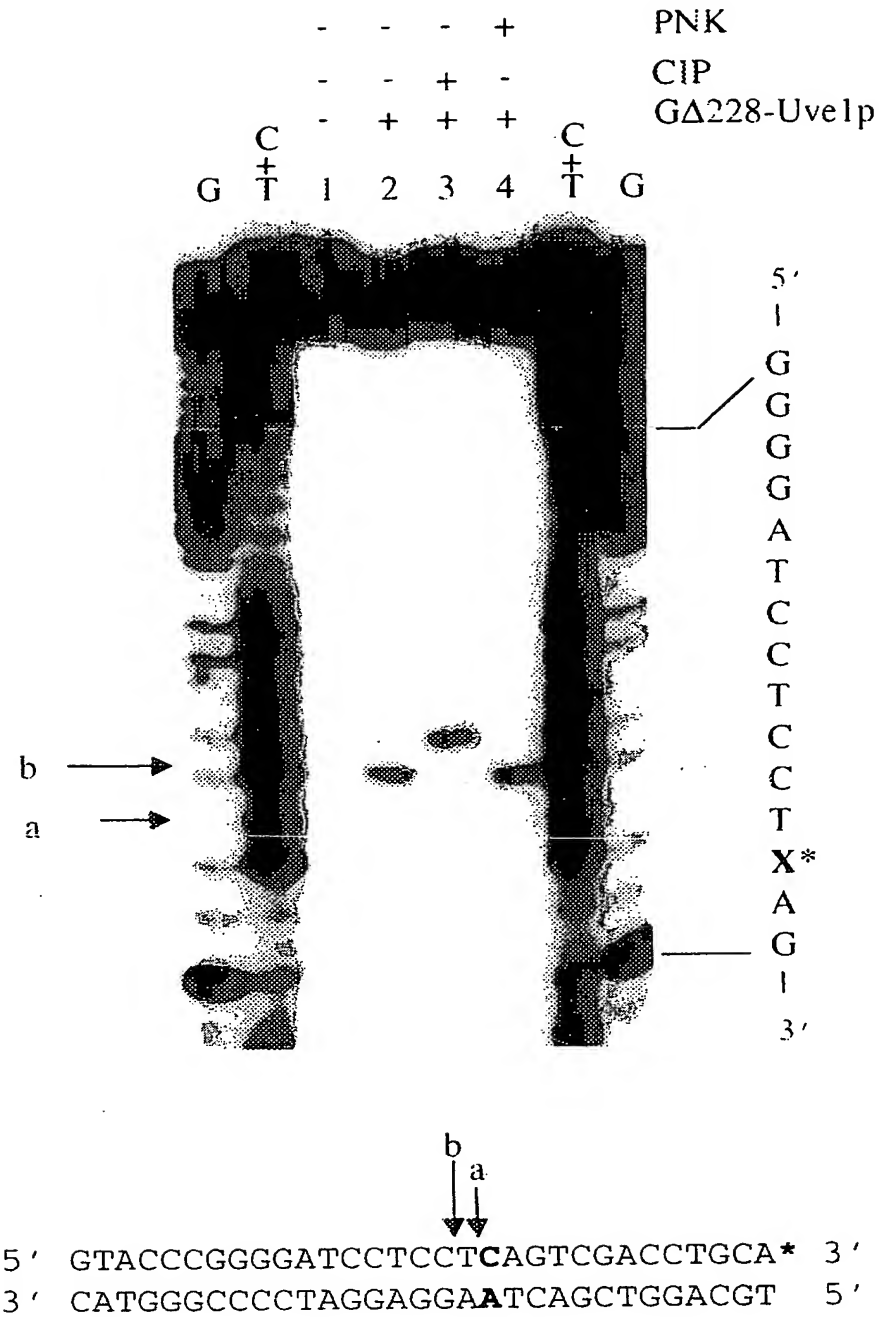
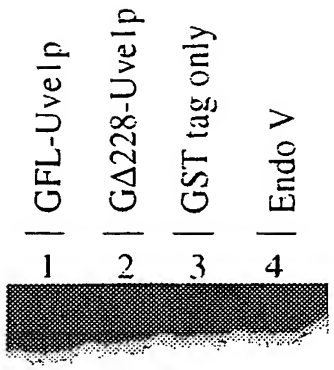
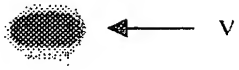
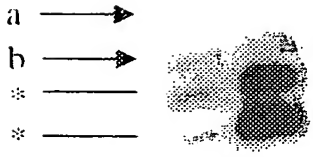


FIG. 11A







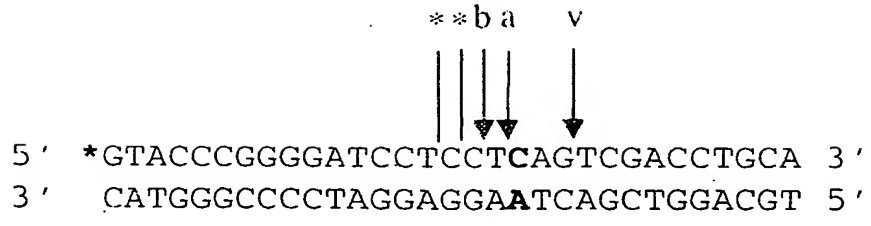


FIG. 11B

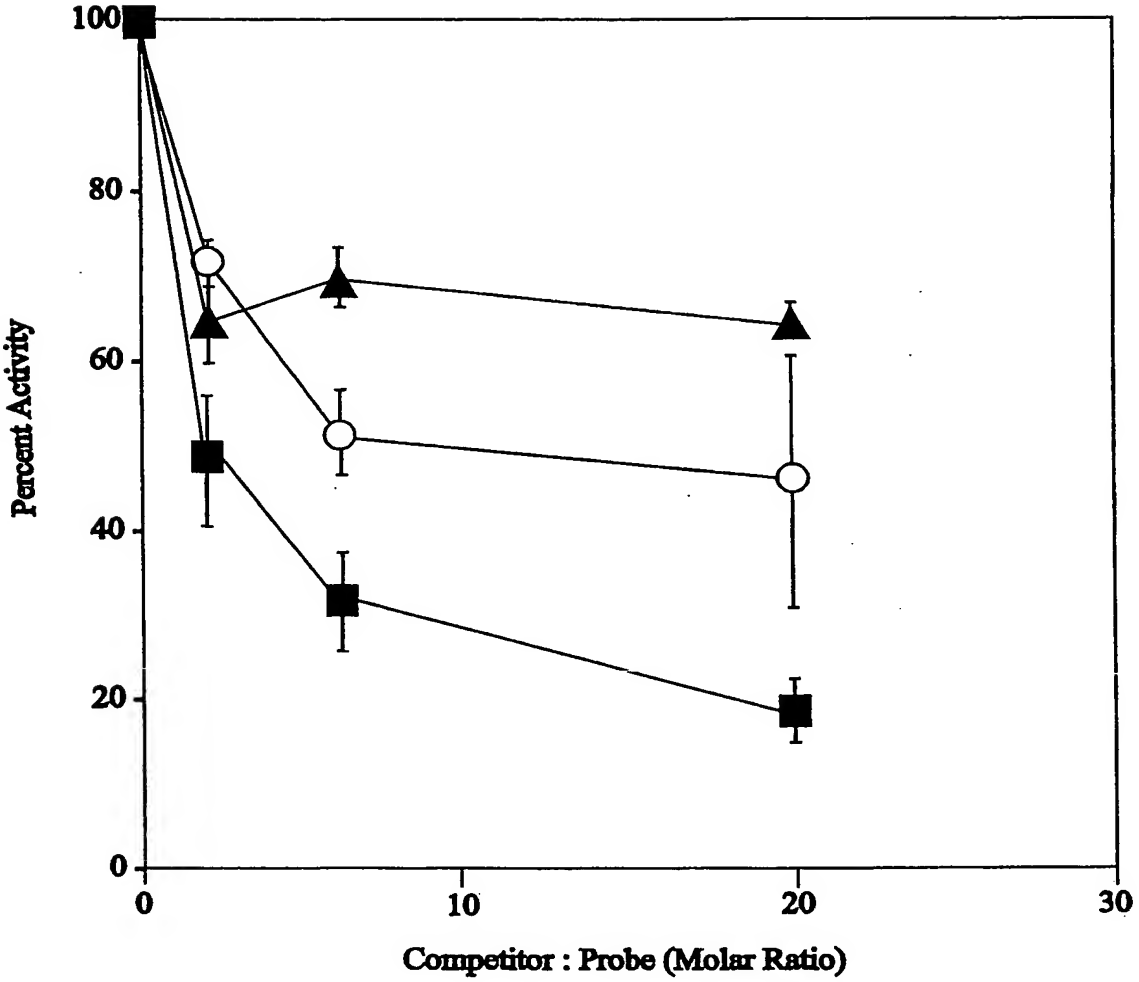


FIG. 12

FIG. 13A

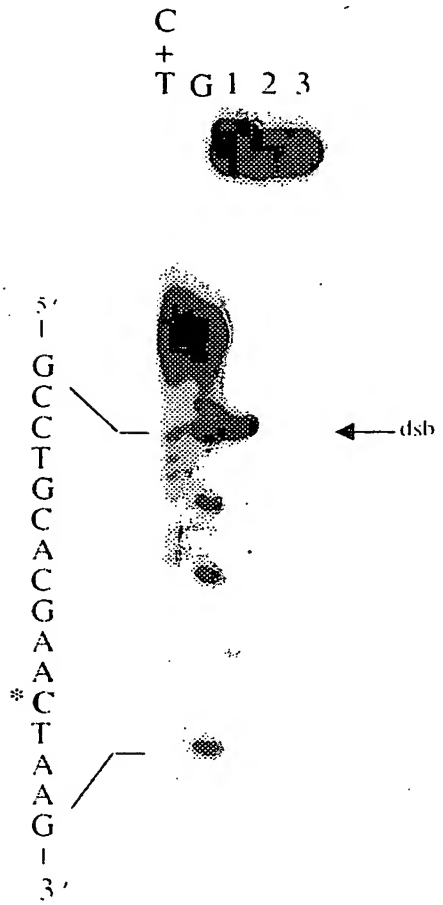
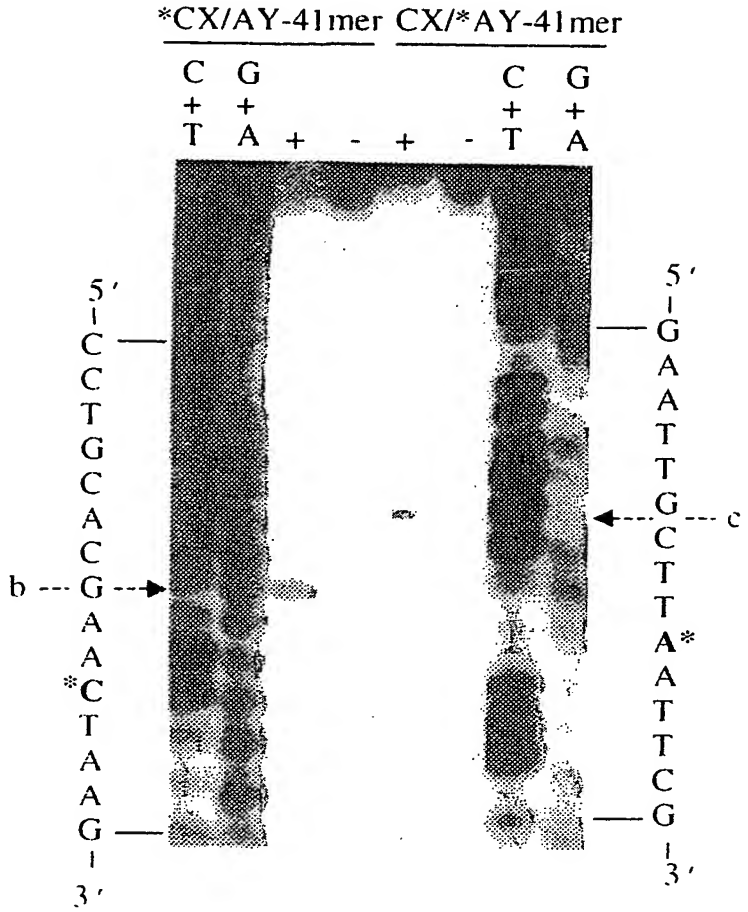
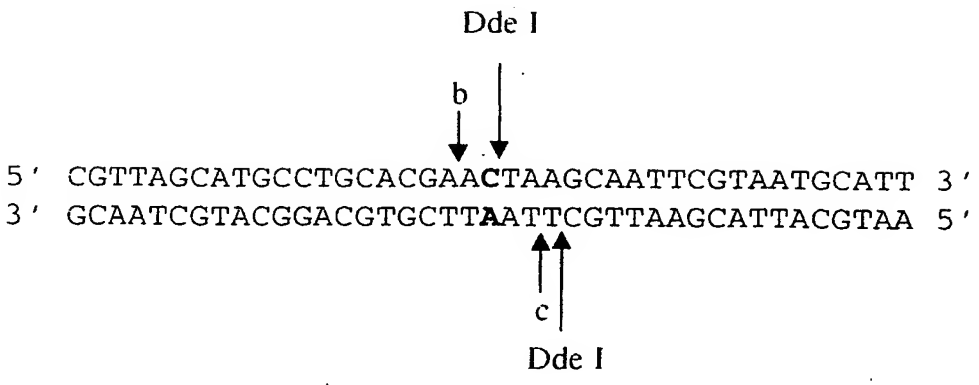


FIG. 13B





SEQUENCE LISTING

<110> EMORY UNIVERSITY

<120> Broad Specificity DNA Damage Endonuclease

<130> 25-98 WO

<140> unassigned

<141> 1999-06-08

<150> US 60/088,521

<151> 1998-06-08

<150> US 60/134,752

<151> 1999-05-18

<160> 39

<170> PatentIn Ver. 2.0

<210> 1

<211> 2492

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:fusion of a GST signal peptide and the UVDE protein of Schizosaccharomyces pombe

<400> 1

atgaccaagt tacctatact aggttattgg aaaaattaag ggccttgtgc aacccactcg 60

acttcttttg gaatatcttg aagaaaaata tgaagagcat ttgtatgagc gcgatgaagg 120

tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc ttccttatta 180

tattgatggt gatgttaaat taacacagtc tatggccatc atacgttata tagctgacaa 240

gcacaacatg ttggttggtt gtccaaaaga gcgtgcagag atttcaatgc ttgaaggagc 300

ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact ttgaaactct 360

caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag atcgtttatg 420

tcataaaaca tatttaaatg ttgaccatgt aacccatcct gacttcatgt tgtatgacgc 480

tcttgatgtt gttttataca tggacccaat gtgcctggat gcgttcccaa aattagtttg 540

ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat ccagcaagta 600

tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc atcctccaaa 660

atcggatcat ctggttccgc gtggatccat gcttaggcta ttgaaacgaa atattcaaat 720

ctctaaacgc attgttttca ccatattaaa acaaaaggca tttaaaggta atcatccttg 780

tgtaccgtcg gtttgtacca ttacttactc tcgttttcat tgtttacccg atacccttaa 840

aagtttactt ccaatgagct caaaaaccac actctcaatg ttaccgcaag ttaatatcgg 900

tgcgaattca ttctctgccg aaacaccagt cgacttaaaa aaagaaaatg agactgagtt 960

agctaatatc agtggacctc acaaaaaaag tacttctacg tctacacgaa agagggcacg 1020

tagcagtaaa aagaaagcga cagattctgt ttccgataaa attgatgagt ctgttgcgtc 1080

ctatgattct tcaactcatc ttaggcgatc gtcgagatca aaaaaaccgg tcaactacaa 1140

ttcctcgtca gaatccgaat cggaggagca aattagtaaa gctactaaaa aagttaaaca 1200

aaaagaggaa gaggagtatg ttgaagaagt cgacgaaaag tctcttaaaa atgaaagtag 1260

ctctgacgag ttcgaaccgg ttgtgccgga acagttggaa actccaattt ctaaacgaag 1320

acggtctcgt tcttctgcaa aaaatttaga aaaagaatct acaatgaatc ttgatgatca 1380

tgctccacga gagatgtttg attgtttgga caaacccata ccctggcgag gacgattggg 1440

gtatgcttgt ttgaatacta ttttaaggtc aatgaaggag agggtttttt gttcacgcac 1500

ctgccgaatt acaaccattc aacgtgatgg gctcgaaagt gtcaagcagc taggtacgca 1560

aaatgtttta gatttaatca aattggttga gtggaatcac aactttggca ttcacttcat 1620

gagagtgagt tctgatttat ttcctttcgc aagccatgca aagtatggat ataccccttga 1680

atttgcacaa tctcatctcg aggaggtggg caagctggca aataaatata atcatcgatt 1740

gactatgcat cctggtcagt acacccagat agcctctcca cgagaagtcg tagttgattc 1800

ggcaatacgt gatttggctt atcatgatga aattctcagt cgtatgaagt tgaatgaaca 1860

attaaataaa gacgctgttt taattattca ccttggtggt acctttgaag gaaaaaaaga 1920

aacattggat aggtttcgta aaaattatca acgcttgtct gattcggtta aagctcgttt 1980

agttttagaa aacgatgatg tttcttggtc agttcaagat ttattacctt tatgccaaga 2040

acttaatatt cctctagttt tggattggca tcatcacaac atagtgccag gaacgcttcg 2100

tgaaggaagt ttagatttaa tgccattaat cccaactatt cgagaaacct ggacaagaaa 2160

gggaattaca cagaagcaac attactcaga atcggctgat ccaacggcga tttctgggat 2220

gaaacgacgt gctcactctg atagggtgtt tgactttcca ccgtgtgatc ctacaatgga 2280

tctaatgata gaagctaagg aaaaggaaca ggctgtattt gaattgtgta gacgttatga 2340

gttacaaaat ccaccatgtc ctcttgaaat tatggggcct gaatacgatc aaactcgaga 2400

tggatattat ccgcccggag ctgaaaagcg tttaactgca agaaaaaggc gtagtagaaa 2460

agaagaagta gaagaggatg aaaaataaaa at 2492

<210> 2

<211> 828

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:fusion protein

(GST leader peptide and Schizosaccharomyces pombe

UVDE

<400> 2

```
Met Thr Lys Leu Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val
 1               5                  10                  15

Gln Pro Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu
            20                  25                  30

His Leu Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe
        35                  40                  45

Glu Leu Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp
    50                  55                  60

Val Lys Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys
65                  70                  75                  80

His Asn Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met
                85                  90                  95
```

Leu Glu Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala

100 105 110

Tyr Ser Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu

115 120 125

Pro Glu Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr

130 135 140

Leu Asn Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala

145 150 155 160

Leu Asp Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro

165 170 175

Lys Leu Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp

180 185 190

Lys Tyr Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp
195 200 205

Gln Ala Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp His Leu
210 215 220

Val Pro Arg Gly Ser Met Leu Arg Leu Leu Lys Arg Asn Ile Gln Ile
225 230 235 240

Ser Lys Arg Ile Val Phe Thr Ile Leu Lys Gln Lys Ala Phe Lys Gly
245 250 255

Asn His Pro Cys Val Pro Ser Val Cys Thr Ile Thr Tyr Ser Arg Phe
260 265 270

His Cys Leu Pro Asp Thr Leu Lys Ser Leu Leu Pro Met Ser Ser Lys
275 280 285

Thr Thr Leu Ser Met Leu Pro Gln Val Asn Ile Gly Ala Asn Ser Phe

290 295 300

Ser Ala Glu Thr Pro Val Asp Leu Lys Lys Glu Asn Glu Thr Glu Leu

305 310 315 320

Ala Asn Ile Ser Gly Pro His Lys Lys Ser Thr Ser Thr Ser Thr Arg

325 330 335

Lys Arg Ala Arg Ser Ser Lys Lys Lys Ala Thr Asp Ser Val Ser Asp

340 345 350

Lys Ile Asp Glu Ser Val Ala Ser Tyr Asp Ser Ser Thr His Leu Arg

355 360 365

Arg Ser Ser Arg Ser Lys Lys Pro Val Asn Tyr Asn Ser Ser Ser Glu

370 375 380

Ser Glu Ser Glu Glu Gln Ile Ser Lys Ala Thr Lys Lys Val Lys Gln

385 390 395 400

Lys Glu Glu Glu Glu Tyr Val Glu Glu Val Asp Glu Lys Ser Leu Lys

405 410 415

Asn Glu Ser Ser Ser Asp Glu Phe Glu Pro Val Val Pro Glu Gln Leu

420 425 430

Glu Thr Pro Ile Ser Lys Arg Arg Arg Ser Arg Ser Ser Ala Lys Asn

435 440 445

Leu Glu Lys Glu Ser Thr Met Asn Leu Asp Asp His Ala Pro Arg Glu

450 455 460

Met Phe Asp Cys Leu Asp Lys Pro Ile Pro Trp Arg Gly Arg Leu Gly

465 470 475 480

Tyr Ala Cys Leu Asn Thr Ile Leu Arg Ser Met Lys Glu Arg Val Phe
485 490 495

Cys Ser Arg Thr Cys Arg Ile Thr Thr Ile Gln Arg Asp Gly Leu Glu
500 505 510

Ser Val Lys Gln Leu Gly Thr Gln Asn Val Leu Asp Leu Ile Lys Leu
515 520 525

Val Glu Trp Asn His Asn Phe Gly Ile His Phe Met Arg Val Ser Ser
530 535 540

Asp Leu Phe Pro Phe Ala Ser His Ala Lys Tyr Gly Tyr Thr Leu Glu
545 550 555 560

Phe Ala Gln Ser His Leu Glu Glu Val Gly Lys Leu Ala Asn Lys Tyr
565 570 575

Asn His Arg Leu Thr Met His Pro Gly Gln Tyr Thr Gln Ile Ala Ser

580 585 590

Pro Arg Glu Val Val Val Asp Ser Ala Ile Arg Asp Leu Ala Tyr His

595 600 605

Asp Glu Ile Leu Ser Arg Met Lys Leu Asn Glu Gln Leu Asn Lys Asp

610 615 620

Ala Val Leu Ile Ile His Leu Gly Gly Thr Phe Glu Gly Lys Lys Glu

625 630 635 640

Thr Leu Asp Arg Phe Arg Lys Asn Tyr Gln Arg Leu Ser Asp Ser Val

645 650 655

Lys Ala Arg Leu Val Leu Glu Asn Asp Asp Val Ser Trp Ser Val Gln

660 665 670

Asp Leu Leu Pro Leu Cys Gln Glu Leu Asn Ile Pro Leu Val Leu Asp

675 680 685

Trp His His His Asn Ile Val Pro Gly Thr Leu Arg Glu Gly Ser Leu

690 695 700

Asp Leu Met Pro Leu Ile Pro Thr Ile Arg Glu Thr Trp Thr Arg Lys

705 710 715 720

Gly Ile Thr Gln Lys Gln His Tyr Ser Glu Ser Ala Asp Pro Thr Ala

725 730 735

Ile Ser Gly Met Lys Arg Arg Ala His Ser Asp Arg Val Phe Asp Phe

740 745 750

Pro Pro Cys Asp Pro Thr Met Asp Leu Met Ile Glu Ala Lys Glu Lys

755 760 765

Glu Gln Ala Val Phe Glu Leu Cys Arg Arg Tyr Glu Leu Gln Asn Pro

770 775 780

Pro Cys Pro Leu Glu Ile Met Gly Pro Glu Tyr Asp Gln Thr Arg Asp

785 790 795 800

Gly Tyr Tyr Pro Pro Gly Ala Glu Lys Arg Leu Thr Ala Arg Lys Arg

805 810 815

Arg Ser Arg Lys Glu Glu Val Glu Glu Asp Glu Lys

820 825

<210> 3

<211> 1161

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:sequence

encoding Schizosaccharomyces pombe UVDE protein

beginning at amino acid 228 of the whole protein

<400> 3

gatgatcatg ctccacgaga gatgtttgat tgtttggaca aacccatacc ctggcgagga 60

cgattggggt atgcttgttt gaatactatt ttaaggtcaa tgaaggagag ggtttttttgt 120

tcacgcacct gccgaattac aaccattcaa cgtgatgggc tcgaaagtgt caagcagcta 180

ggtacgcaaa atgttttaga tttaatcaaa ttggttgagt ggaatcacaa ctttggcatt 240

cacttcatga gagtgagttc tgatttattt cctttcgcaa gccatgcaaa gtatggatat 300

acccttgaat ttgcacaatc tcatctcgag gaggtgggca agctggcaaa taaatataat 360

catcgattga ctatgcatcc tggtcagtac acccagatag cctctccacg agaagtcgta 420

gttgattcgg caatacgtga tttggcttat catgatgaaa ttctcagtcg tatgaagttg 480

aatgaacaat taaataaaga cgctgtttta attattcacc ttggtggtac ctttgaagga 540

aaaaaagaaa cattggatag gtttcgtaaa aattatcaac gcttgtctga ttcggttaaa 600

gctcgtttag ttttagaaaa cgatgatgtt tcttggtcag ttcaagattt attaccttta 660

tgccaagaac ttaatattcc tctagttttg gattggcatc atcacaacat agtgccagga 720

acgcttcgtg aaggaagttt agatttaatg ccattaatcc caactattcg agaaacctgg 780

acaagaaagg gaattacaca gaagcaacat tactcagaat cggctgatcc aacggcgatt 840

tctgggatga aacgacgtgc tcactctgat agggtgtttg actttccacc gtgtgatcct 900

acaatggatc taatgataga agctaaggaa aaggaacagg ctgtatttga attgtgtaga 960

cgttatgagt tacaaaatcc accatgtcct cttgaaatta tggggcctga atacgatcaa 1020

actcgagatg gatattatcc gcccggagct gaaaagcgtt taactgcaag aaaaaggcgt 1080

agtagaaaag aagaagtaga agaggatgaa aaataaaaat ccgtcatact ttttgattta 1140

tggcataatt tagccatctc c 1161

<210> 4

<211> 371

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:truncated

derivative of the S. pombe UVDE protein

<400> 4

Asp Asp His Ala Pro Arg Glu Met Phe Asp Cys Leu Asp Lys Pro Ile
1 5 10 15

Pro Trp Arg Gly Arg Leu Gly Tyr Ala Cys Leu Asn Thr Ile Leu Arg
20 25 30

Ser Met Lys Glu Arg Val Phe Cys Ser Arg Thr Cys Arg Ile Thr Thr
35 40 45

Ile Gln Arg Asp Gly Leu Glu Ser Val Lys Gln Leu Gly Thr Gln Asn
50 55 60

Val Leu Asp Leu Ile Lys Leu Val Glu Trp Asn His Asn Phe Gly Ile
65 70 75 80

His Phe Met Arg Val Ser Ser Asp Leu Phe Pro Phe Ala Ser His Ala
85 90 95

Lys Tyr Gly Tyr Thr Leu Glu Phe Ala Gln Ser His Leu Glu Glu Val

100 105 110

Gly Lys Leu Ala Asn Lys Tyr Asn His Arg Leu Thr Met His Pro Gly

115 120 125

Gln Tyr Thr Gln Ile Ala Ser Pro Arg Glu Val Val Val Asp Ser Ala

130 135 140

Ile Arg Asp Leu Ala Tyr His Asp Glu Ile Leu Ser Arg Met Lys Leu

145 150 155 160

Asn Glu Gln Leu Asn Lys Asp Ala Val Leu Ile Ile His Leu Gly Gly

165 170 175

Thr Phe Glu Gly Lys Lys Glu Thr Leu Asp Arg Phe Arg Lys Asn Tyr

180 185 190

Gln Arg Leu Ser Asp Ser Val Lys Ala Arg Leu Val Leu Glu Asn Asp

195 200 205

Asp Val Ser Trp Ser Val Gln Asp Leu Leu Pro Leu Cys Gln Glu Leu

210 215 220

Asn Ile Pro Leu Val Leu Asp Trp His His His Asn Ile Val Pro Gly

225 230 235 240

Thr Leu Arg Glu Gly Ser Leu Asp Leu Met Pro Leu Ile Pro Thr Ile

245 250 255

Arg Glu Thr Trp Thr Arg Lys Gly Ile Thr Gln Lys Gln His Tyr Ser

260 265 270

Glu Ser Ala Asp Pro Thr Ala Ile Ser Gly Met Lys Arg Arg Ala His

275 280 285

Ser Asp Arg Val Phe Asp Phe Pro Pro Cys Asp Pro Thr Met Asp Leu
290 295 300

Met Ile Glu Ala Lys Glu Lys Glu Gln Ala Val Phe Glu Leu Cys Arg
305 310 315 320

Arg Tyr Glu Leu Gln Asn Pro Pro Cys Pro Leu Glu Ile Met Gly Pro
325 330 335

Glu Tyr Asp Gln Thr Arg Asp Gly Tyr Tyr Pro Pro Gly Ala Glu Lys
340 345 350

Arg Leu Thr Ala Arg Lys Arg Arg Ser Arg Lys Glu Glu Val Glu Glu
355 360 365

Asp Glu Lys
370

<210> 5

<211> 1811

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:nucleotide sequence encoding fusion protein (GST signal peptide fused to the truncated derivative of the S. pmbe UVDE protein)

<400> 5

atgaccaagt tacctatact aggttattgg aaaaattaag ggccttgtgc aacccactcg 60

acttcttttg gaatatcttg aagaaaaata tgaagagcat ttgtatgagc gcgatgaagg 120

tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc ttccttatta 180

tattgatggt gatgttaaat taacacagtc tatggccatc atacgttata tagctgacaa 240

gcacaacatg ttggttggtt gtccaaaaga gcgtgcagag atttcaatgc ttgaaggagc 300

ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact ttgaaactct 360

caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag atcgtttatg 420

tcataaaaca tatttaaatg ttgaccatgt aacccatcct gacttcatgt tgtatgacgc 480

tcttgatgtt gttttataca tggacccaat gtgcctggat gcgttcccaa aattagtttg 540

ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat ccagcaagta 600

tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc atcctccaaa 660

atcggatcat ctggttccgc gtggatccga tgatcatgct ccacgagaga tgtttgattg 720

tttggacaaa cccataccct ggcgaggacg attggggtat gcttgtttga atactatttt 780

aaggtcaatg aaggagaggg ttttttgttc acgcacctgc cgaattacaa ccattcaacg 840

tgatgggctc gaaagtgtca agcagctagg tacgcaaaat gttttagatt taatcaaatt 900

ggttgagtgg aatcacaact ttggcattca cttcatgaga gtgagttctg atttatttcc 960

tttcgcaagc catgcaaagt atggatatac ccttgaattt gcacaatctc atctcgagga 1020

ggtgggcaag ctggcaaata aatataatca tcgattgact atgcatcctg gtcagtacac 1080

ccagatagcc tctccacgag aagtcgtagt tgattcggca atacgtgatt tggcttatca 1140

tgatgaaatt ctcagtcgta tgaagttgaa tgaacaatta aataaagacg ctgttttaat 1200

tattcacctt ggtggtacct ttgaaggaaa aaaagaaaca ttggataggt ttcgtaaaaa 1260

ttatcaacgc ttgtctgatt cggttaaagc tcgtttagtt ttagaaaacg atgatgtttc 1320

ttggtcagtt caagatttat tacctttatg ccaagaactt aatattcctc tagttttgga 1380

ttggcatcat cacaacatag tgccaggaac gcttcgtgaa ggaagtttag atttaatgcc 1440

attaatccca actattcgag aaacctggac aagaaaggga attacacaga agcaacatta 1500

ctcagaatcg gctgatccaa cggcgatttc tgggatgaaa cgacgtgctc actctgatag 1560

ggtgtttgac tttccaccgt gtgatcctac aatggatcta atgatagaag ctaaggaaaa 1620

ggaacaggct gtatttgaat tgtgtagacg ttatgagtta caaaatccac catgtcctct 1680

tgaaattatg gggcctgaat acgatcaaac tcgagatgga tattatccgc ccggagctga 1740

aaagcgttta actgcaagaa aaaggcgtag tagaaaagaa gaagtagaag aggatgaaaa 1800

ataaggatcc c 1811

<210> 6

<211> 600

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:fusion protein

comprising the GST signal peptide and the

truncated UVDE protein of S. pombe

<400> 6

Met Thr Lys Leu Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val

1 5 10 15

Gln Pro Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu

20 25 30

His Leu Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe

35 40 45

Glu Leu Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp

50 55 60

Val Lys Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys

65 70 75 80

His Asn Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met

85 90 95

Leu Glu Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala

100 105 110

Tyr Ser Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu

115 120 125

Pro Glu Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr

130 135 140

Leu Asn Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala

145 150 155 160

Leu Asp Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro

165 170 175

Lys Leu Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp
180 185 190

Lys Tyr Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp
195 200 205

Gln Ala Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp His Leu
210 215 220

Val Pro Arg Gly Ser Asp Asp His Ala Pro Arg Glu Met Phe Asp Cys
225 230 235 240

Leu Asp Lys Pro Ile Pro Trp Arg Gly Arg Leu Gly Tyr Ala Cys Leu
245 250 255

Asn Thr Ile Leu Arg Ser Met Lys Glu Arg Val Phe Cys Ser Arg Thr
260 265 270

Cys Arg Ile Thr Thr Ile Gln Arg Asp Gly Leu Glu Ser Val Lys Gln

275 280 285

Leu Gly Thr Gln Asn Val Leu Asp Leu Ile Lys Leu Val Glu Trp Asn

290 295 300

His Asn Phe Gly Ile His Phe Met Arg Val Ser Ser Asp Leu Phe Pro

305 310 315 320

Phe Ala Ser His Ala Lys Tyr Gly Tyr Thr Leu Glu Phe Ala Gln Ser

325 330 335

His Leu Glu Glu Val Gly Lys Leu Ala Asn Lys Tyr Asn His Arg Leu

340 345 350

Thr Met His Pro Gly Gln Tyr Thr Gln Ile Ala Ser Pro Arg Glu Val

355 360 365

Val Val Asp Ser Ala Ile Arg Asp Leu Ala Tyr His Asp Glu Ile Leu

370 375 380

Ser Arg Met Lys Leu Asn Glu Gln Leu Asn Lys Asp Ala Val Leu Ile

385 390 395 400

Ile His Leu Gly Gly Thr Phe Glu Gly Lys Lys Glu Thr Leu Asp Arg

405 410 415

Phe Arg Lys Asn Tyr Gln Arg Leu Ser Asp Ser Val Lys Ala Arg Leu

420 425 430

Val Leu Glu Asn Asp Asp Val Ser Trp Ser Val Gln Asp Leu Leu Pro

435 440 445

Leu Cys Gln Glu Leu Asn Ile Pro Leu Val Leu Asp Trp His His His

450 455 460

Asn Ile Val Pro Gly Thr Leu Arg Glu Gly Ser Leu Asp Leu Met Pro
465 470 475 480

Leu Ile Pro Thr Ile Arg Glu Thr Trp Thr Arg Lys Gly Ile Thr Gln
485 490 495

Lys Gln His Tyr Ser Glu Ser Ala Asp Pro Thr Ala Ile Ser Gly Met
500 505 510

Lys Arg Arg Ala His Ser Asp Arg Val Phe Asp Phe Pro Pro Cys Asp
515 520 525

Pro Thr Met Asp Leu Met Ile Glu Ala Lys Glu Lys Glu Gln Ala Val
530 535 540

Phe Glu Leu Cys Arg Arg Tyr Glu Leu Gln Asn Pro Pro Cys Pro Leu
545 550 555 560

Glu Ile Met Gly Pro Glu Tyr Asp Gln Thr Arg Asp Gly Tyr Tyr Pro

565 570 575

Pro Gly Ala Glu Lys Arg Leu Thr Ala Arg Lys Arg Arg Ser Arg Lys

580 585 590

Glu Glu Val Glu Glu Asp Glu Lys

595 600

<210> 7

<211> 688

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:nucleotide

sequence encoding GST signal (leader) peptide

<400> 7

atgaccaagt tacctatact aggttattgg aaaaattaag ggccttgtgc aacccactcg 60

acttcttttg gaatatcttg aagaaaaata tgaagagcat ttgtatgagc gcgatgaagg 120

tgataaatgg cgaaacaaaa agtttgaatt gggtttggag tttcccaatc ttccttatta 180

tattgatggt gatgttaaat taacacagtc tatggccatc atacgttata tagctgacaa 240

gcacaacatg ttggttggtt gtccaaaaga gcgtgcagag atttcaatgc ttgaaggagc 300

ggttttggat attagatacg gtgtttcgag aattgcatat agtaaagact ttgaaactct 360

caaagttgat tttcttagca agctacctga aatgctgaaa atgttcgaag atcgtttatg 420

tcataaaaca tatttaaatg ttgaccatgt aacccatcct gacttcatgt tgtatgacgc 480

tcttgatgtt gttttataca tggacccaat gtgcctggat gcgttcccaa aattagtttg 540

ttttaaaaaa cgtattgaag ctatcccaca aattgataag tacttgaaat ccagcaagta 600

tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc atcctccaaa 660

atcggatcat ctggttccgc gtggatcc 688

<210> 8

<211> 229

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:amino acid sequence of the GST leader peptide

<400> 8

Met Thr Lys Leu Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val
1 5 10 15

Gln Pro Thr Arg Leu Leu Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu
20 25 30

His Leu Tyr Glu Arg Asp Glu Gly Asp Lys Trp Arg Asn Lys Lys Phe
35 40 45

Glu Leu Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp Gly Asp
50 55 60

Val Lys Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys

65 70 75 80

His Asn Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met

85 90 95

Leu Glu Gly Ala Val Leu Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala

100 105 110

Tyr Ser Lys Asp Phe Glu Thr Leu Lys Val Asp Phe Leu Ser Lys Leu

115 120 125

Pro Glu Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys Thr Tyr

130 135 140

Leu Asn Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala

145 150 155 160

Leu Asp Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro

165 170 175

Lys Leu Val Cys Phe Lys Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp

180 185 190

Lys Tyr Leu Lys Ser Ser Lys Tyr Ile Ala Trp Pro Leu Gln Gly Trp

195 200 205

Gln Ala Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp His Leu

210 215 220

Val Pro Arg Gly Ser

225

<210> 9

<211> 36

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:oligonucleotide used in polymerase chain reaction.

<400> 9

<210> 10

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:oligonucleotide used in polymerase chain reaction.

<400> 10

ggccatggtt atttttcatc ctc 23

<210> 11

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:oligonucleotide used in polymerase chain reaction

<400> 11

aatgggatcc gatgatcatg ctccacga 28

<210> 12

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:oligonucleotide used in polymerase chain reaction

<400> 12

gggatcctta tttttcatcc tcttctac 28

<210> 13

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded

oligonucleotide containing cis-syn cyclobutane pyrimidine dimer

<220>

<221> misc_feature

<222> (15)..(16)

<223> At positions 15-16, the T-T is in the form of a cis-syn cyclobutane pyrimidine dimer

<400> 13

catgcctgca cgaattaagc aattcgtaat 30

<210> 14

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:undamaged
double stranded oligonucleotide

<400> 14

catgcctgca cgaattaagc aattcgtaat 30

<210> 15

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded
oligonucleotide containing cis-syn cyclobutane
dimer at positions 21-22

<220>

<221> misc_feature

<222> (21)..(22)

<223> T-T at positions 21-22 are in the form of a

cis-syn cyclobutane pyrimidine dimer.

<400> 15

agctaccatg cctgcacgaa ttaagcaatt cgtaatcatg gtcatagct 49

<210> 16

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded

oligonucleotide containing cis-syn cyclobutane

pyrimidine dimer at positions 21-22.

<220>

<221> misc_feature

<222> (21)..(22)

<223> T-T at positions 21-22 is in the form of a

trans-syn I cyclobutane pyrimidine dimer

<400> 16

agctaccatg cctgcacgaa ttaagcaatt cgtaatcatg gtcatagct 49

<210> 17

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide containing trans-syn II cyclobutane pyrimidine dimer at positions 21-22

<220>

<221> misc_feature

<222> (21)..(22)

<223> T-T at positions 21-22 is in the form of a trans-syn II cyclobutane pyrimidine dimer.

<400> 17

agctaccatg cctgcacgaa ttaagcaatt cgtaatcatg gtcatagct 49

<210> 18

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide containing a 6-4 photoproduct

<220>

<221> misc_feature

<222> (21)..(22)

<223> T-T at positions 21-22 is in the form of a 6-4 photoproduct

<400> 18

agctaccatg cctgcacgaa ttaagcaatt cgtaatcatg gtcatagct 49

<210> 19

<211> 49

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide containing Dewar isomer

<220>

<221> misc_feature

<222> (21)..(22)

<223> T-T at positions 21-22 is in the form of a Dewar isomer

<400> 19

agctaccatg cctgcacgaa ttaagcaatt cgtaatcatg gtcatagct 49

<210> 20

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide containing cisplatin DNA diadduct

<220>

<221> misc_feature

<222> (16)..(17)

<223> G-G at positions 16-17 are in the form of a platinum-DNA diadduct

<400> 20

tccctccttc cttccggccc tccttcccct tc 32

<210> 21

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein n is uracil

<400> 21

cttggactgg atgtcggcac nagcggatac aggagca 37

<210> 22

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein n is dihydrouracil

<220>

<221> misc_feature

<222> (21)

<223> At position 21, n is dihydrouracil

<400> 22

cttggactgg atgtcggcac nagcggatac aggagca 37

<210> 23

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein n at postion 21 represents an abasic site

<220>

<221> misc_feature

<222> (21)

<223> Position 21 (n) is an abasic site

<400> 23

cttggactgg atgtcggcac nagcggatac aggagca 37

<210> 24

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein n at position 13 is an inosine

<220>

<221> misc_feature

<222> (13)

<223> N at position 13 is inosine

<400> 24

tgcaggtcga ctnaggagga tccccgggta c 31

<210> 25

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded
oligonucleotide wherein n at position 13
represents xanthine

<220>

<221> misc_feature

<222> (13)

<223> N at position 13 represents xanthine

<400> 25

tgcaggtcga ctnaggagga tccccgggta c 31

<210> 26

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein position 21 is 8-oxoguanine

<220>

<221> misc_feature

<222> (21)

<223> N at position 21 is 8-oxoguanine

<400> 26

cttggactgg atgtcggcac nagcggatac aggagca 37

<210> 27

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide representing all 16 possible base pair mismatches at position 18 in individual preparations

<220>

<221> misc_feature

<222> (18)

<223> This entry represents preparations (16)
containing all possible mispairing at position 18

<400> 27

gtacccgggg atcctccnag tcgacctgca 30

<210> 28

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded
oligonucleotide containing a CA mismatched base
pair at position 21

<220>

<221> misc_feature

<222> (21)

<223> At position 21 n represents a C/A mismatched base pair in the double stranded oligonucleotide

<400> 28

cgttagcatg cctgcacgaa ntaagcaatt cgtaatgcat t 41

<210> 29

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein there is a C/A mismatched base pair at position 36

<220>

<221> misc_feature

<222> (36)

<223> At position 36, n respresents a C/A mismatched

base pair (C on the given strand)

<400> 29

cgttacaagt ccgtcacgaa ttaagcaatt cgtaangcat t 41

<210> 30

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded

oligonucleotide wherein position 31 is a C/A

mimatched base pair

<220>

<221> misc_feature

<222> (31)

<223> The n at position 31 represents C of a C/A mismatched base pair

<400> 30

cgttacaagt ccgtcacgaa ttaagcaatt ngtaacgcat t 41

<210> 31

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:doublestranded oligonucleotide wherein n at position 26 represents a C/A mismatched base pair

<220>

<221> misc_feature

<222> (26)

<400> 31

cgttacaagt ccgtcacgaa ttaagnaatt cgtaacgcat t 41

<210> 32

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein there is a C/A mismatched base pair at position 20

<220>

<221> misc_feature

<222> (20)

<223> The n at position 21 represents a c/a mismatched base pair, with the c within the given sequence

<400> 32

cgttacaagt ccgtcacgac ttaagcaatt cgtaacgcat t 41

<210> 33

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein n at position 15 represents a C/A mismatched base pair

<220>

<221> misc_feature

<222> (15)

<223> N at position 15 is a C/A mismatched base pair (C on the given strand)

<400> 33

cgttacaagt ccgtnacgaa ttaagcaatt cgtaacgcat t 41

<210> 34

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein n at position 10 is a C/A mismatched base pair

<220>

<221> misc_feature

<222> (10)

<223> N at position 10 is a C/A mismatched base pair (C on the given strand)

<400> 34

cgttacaagn ccgtcacgaa ttaagcaatt cgtaacgcat t 41

<210> 35

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:double stranded oligonucleotide wherein n at position 5 is a C/A mismatched base pair

<220>

<221> misc_feature

<222> (5)

<223> N at position 5 represents a C/A mismatched base pair (C on the given strand)

<400> 35

cgttncaagt ccgtcacgaa ttaagcaatt cgtaacgcat t 41

<210> 36

<211> 658

<212> PRT

<213> Neurospora crassa

<400> 36

Met Pro Ser Arg Lys Ser Lys Ala Ala Ala Leu Asp Thr Pro Gln Ser

1 5 10 15

Glu Ser Ser Thr Phe Ser Ser Thr Leu Asp Ser Ser Ala Pro Ser Pro

20 25 30

Ala Arg Asn Leu Arg Arg Ser Gly Arg Asn Ile Leu Gln Pro Ser Ser

35 40 45

Glu Lys Asp Arg Asp His Glu Lys Arg Ser Gly Glu Glu Leu Ala Gly

50 55 60

Arg Met Met Gly Lys Asp Ala Asn Gly His Cys Leu Arg Glu Gly Lys

65 70 75 80

Glu Gln Glu Glu Gly Val Lys Met Ala Ile Glu Gly Leu Ala Arg Met

85 90 95

Glu Arg Arg Leu Gln Arg Ala Thr Lys Arg Gln Lys Lys Gln Leu Glu

100 105 110

Glu Asp Gly Ile Pro Val Pro Ser Val Val Ser Arg Phe Pro Thr Ala

115 120 125

Pro Tyr His His Lys Ser Thr Asn Ala Glu Glu Arg Glu Ala Lys Glu

130 135 140

Pro Val Leu Lys Thr His Ser Lys Asp Val Glu Arg Glu Ala Glu Ile

145 150 155 160

Gly Val Asp Asp Val Val Lys Met Glu Pro Ala Ala Thr Asn Ile Ile

165 170 175

Glu Pro Glu Asp Ala Gln Asp Ala Ala Glu Arg Gly Ala Ala Arg Pro

180 185 190

Pro Ala Val Asn Ser Ser Tyr Leu Pro Leu Pro Trp Lys Gly Arg Leu

195 200 205

Gly Tyr Ala Cys Leu Asn Thr Tyr Leu Arg Asn Ala Lys Pro Pro Ile
210 215 220

Phe Ser Ser Arg Thr Cys Arg Met Ala Ser Ile Val Asp His Arg His
225 230 235 240

Pro Leu Gln Phe Glu Asp Glu Pro Glu His His Leu Lys Asn Lys Pro
245 250 255

Asp Lys Ser Lys Glu Pro Gln Asp Glu Leu Gly His Lys Phe Val Gln
260 265 270

Glu Leu Gly Leu Ala Asn Ala Arg Asp Ile Val Lys Met Leu Cys Trp
275 280 285

Phe Pro Phe Ala Ser His Pro Val His Gly Tyr Lys Leu Ala Pro Phe
290 295 300

Ala Ser Glu Val Leu Ala Glu Ala Gly Arg Val Ala Ala Glu Leu Gly

305 310 315 320

His Arg Leu Thr Thr His Pro Gly Gln Phe Thr Gln Leu Gly Ser Pro

325 330 335

Arg Lys Glu Val Val Glu Ser Ala Ile Arg Asp Leu Glu Tyr His Asp

340 345 350

Glu Leu Leu Ser Leu Leu Lys Leu Pro Glu Gln Gln Asn Arg Asp Ala

355 360 365

Val Met Ile Ile His Met Gly Gly Gln Phe Gly Asp Lys Ala Ala Thr

370 375 380

Leu Glu Arg Phe Lys Arg Asn Tyr Ala Arg Leu Ser Gln Ser Cys Lys

385 390 395 400

Asn Arg Leu Val Leu Glu Asn Asp Asp Val Gly Trp Thr Val His Asp

405 410 415

Leu Leu Pro Val Cys Glu Glu Leu Asn Ile Pro Met Val Leu Asp Tyr

420 425 430

His His His Asn Ile Cys Phe Asp Pro Ala His Leu Arg Glu Gly Thr

435 440 445

Leu Asp Ile Ser Asp Pro Lys Leu Gln Glu Arg Ile Ala Asn Thr Trp

450 455 460

Lys Arg Lys Gly Ile Lys Gln Lys Met His Tyr Ser Glu Pro Cys Asp

465 470 475 480

Gly Ala Val Thr Pro Arg Asp Arg Arg Lys His Arg Pro Arg Val Met

485 490 495

Thr Leu Pro Pro Cys Pro Pro Asp Met Asp Leu Met Ile Glu Ala Lys

500 505 510

Asp Lys Glu Gln Ala Val Phe Glu Leu Met Arg Thr Phe Lys Leu Pro

515 520 525

Gly Phe Glu Lys Ile Asn Asp Met Val Pro Tyr Asp Arg Asp Asp Glu

530 535 540

Asn Arg Pro Ala Pro Pro Val Lys Ala Pro Lys Lys Lys Lys Gly Gly

545 550 555 560

Lys Arg Lys Arg Thr Thr Asp Glu Glu Ala Ala Glu Pro Glu Glu Val

565 570 575

Glu Val Pro Glu Glu Glu Arg Ala Met Gly Gly Pro Tyr Asn Arg Val

580 585 590

Tyr Trp Pro Leu Gly Cys Glu Glu Trp Leu Lys Pro Lys Lys Arg Glu

595 600 605

Val Lys Lys Gly Lys Val Pro Glu Glu Val Glu Asp Glu Gly Glu Phe

610 615 620

Asp Gly

625

<210> 37

<211> 658

<212> PRT

<213> Bacillus subtilis

<400> 37

Met Pro Ser Arg Lys Ser Lys Ala Ala Ala Leu Asp Thr Pro Gln Ser

1 5 10 15

Glu Ser Ser Thr Phe Ser Ser Thr Leu Asp Ser Ser Ala Pro Ser Pro

20 25 30

Ala Arg Asn Leu Arg Arg Ser Gly Arg Asn Ile Leu Gln Pro Ser Ser

35 40 45

Glu Lys Asp Arg Asp His Glu Lys Arg Ser Gly Glu Glu Leu Ala Gly

50 55 60

Arg Met Met Gly Lys Asp Ala Asn Gly His Cys Leu Arg Glu Gly Lys

65 70 75 80

Glu Gln Glu Glu Gly Val Lys Met Ala Ile Glu Gly Leu Ala Arg Met

85 90 95

Glu Arg Arg Leu Gln Arg Ala Thr Lys Arg Gln Lys Lys Gln Leu Glu

100 105 110

Glu Asp Gly Ile Pro Val Pro Ser Val Val Ser Arg Phe Pro Thr Ala
115 120 125

Pro Tyr His His Lys Ser Thr Asn Ala Glu Glu Arg Glu Ala Lys Glu
130 135 140

Pro Val Leu Lys Thr His Ser Lys Asp Val Glu Arg Glu Ala Glu Ile
145 150 155 160

Gly Val Asp Asp Val Val Lys Met Glu Pro Ala Ala Thr Asn Ile Ile
165 170 175

Glu Pro Glu Asp Ala Gln Asp Ala Ala Glu Arg Gly Ala Ala Arg Pro
180 185 190

Pro Ala Val Asn Ser Ser Tyr Leu Pro Leu Pro Trp Lys Gly Arg Leu
195 200 205

Gly Tyr Ala Cys Leu Asn Thr Tyr Leu Arg Asn Ala Lys Pro Pro Ile

210 215 220

Phe Ser Ser Arg Thr Cys Arg Met Ala Ser Ile Val Asp His Arg His

225 230 235 240

Pro Leu Gln Phe Glu Asp Glu Pro Glu His His Leu Lys Asn Lys Pro

245 250 255

Asp Lys Ser Lys Glu Pro Gln Asp Glu Leu Gly His Lys Phe Val Gln

260 265 270

Glu Leu Gly Leu Ala Asn Ala Arg Asp Ile Val Lys Met Leu Cys Trp

275 280 285

Phe Pro Phe Ala Ser His Pro Val His Gly Tyr Lys Leu Ala Pro Phe

290 295 300

Ala Ser Glu Val Leu Ala Glu Ala Gly Arg Val Ala Ala Glu Leu Gly

305 310 315 320

His Arg Leu Thr Thr His Pro Gly Gln Phe Thr Gln Leu Gly Ser Pro

325 330 335

Arg Lys Glu Val Val Glu Ser Ala Ile Arg Asp Leu Glu Tyr His Asp

340 345 350

Glu Leu Leu Ser Leu Leu Lys Leu Pro Glu Gln Gln Asn Arg Asp Ala

355 360 365

Val Met Ile Ile His Met Gly Gly Gln Phe Gly Asp Lys Ala Ala Thr

370 375 380

Leu Glu Arg Phe Lys Arg Asn Tyr Ala Arg Leu Ser Gln Ser Cys Lys

385 390 395 400

Asn Arg Leu Val Leu Glu Asn Asp Asp Val Gly Trp Thr Val His Asp

405 410 415

Leu Leu Pro Val Cys Glu Glu Leu Asn Ile Pro Met Val Leu Asp Tyr

420 425 430

His His His Asn Ile Cys Phe Asp Pro Ala His Leu Arg Glu Gly Thr

435 440 445

Leu Asp Ile Ser Asp Pro Lys Leu Gln Glu Arg Ile Ala Asn Thr Trp

450 455 460

Lys Arg Lys Gly Ile Lys Gln Lys Met His Tyr Ser Glu Pro Cys Asp

465 470 475 480

Gly Ala Val Thr Pro Arg Asp Arg Arg Lys His Arg Pro Arg Val Met

485 490 495

Thr Leu Pro Pro Cys Pro Pro Asp Met Asp Leu Met Ile Glu Ala Lys
500 505 510

Asp Lys Glu Gln Ala Val Phe Glu Leu Met Arg Thr Phe Lys Leu Pro
515 520 525

Gly Phe Glu Lys Ile Asn Asp Met Val Pro Tyr Asp Arg Asp Asp Glu
530 535 540

Asn Arg Pro Ala Pro Pro Val Lys Ala Pro Lys Lys Lys Lys Gly Gly
545 550 555 560

Lys Arg Lys Arg Thr Thr Asp Glu Glu Ala Ala Glu Pro Glu Glu Val
565 570 575

Glu Val Pro Glu Glu Glu Arg Ala Met Gly Gly Pro Tyr Asn Arg Val
580 585 590

Tyr Trp Pro Leu Gly Cys Glu Glu Trp Leu Lys Pro Lys Lys Arg Glu

595 600 605

Val Lys Lys Gly Lys Val Pro Glu Glu Val Glu Asp Glu Gly Glu Phe

610 615 620

Asp Gly

625

<210> 38

<211> 581

<212> PRT

<213> Homo sapiens

<400> 38

Met Gly Thr Thr Gly Leu Glu Ser Leu Ser Leu Gly Asp Arg Gly Ala

1 5 10 15

Ala Pro Thr Val Thr Ser Ser Glu Arg Leu Val Pro Asp Pro Pro Asn

20 25 30

Asp Leu Arg Lys Glu Asp Val Ala Met Glu Leu Glu Arg Val Gly Glu

35 40 45

Asp Glu Glu Gln Met Met Ile Lys Arg Ser Ser Glu Cys Asn Pro Leu

50 55 60

Leu Gln Glu Pro Ile Ala Ser Ala Gln Phe Gly Ala Thr Ala Gly Thr

65 70 75 80

Glu Cys Arg Lys Ser Val Pro Cys Gly Trp Glu Arg Val Val Lys Gln

85 90 95

Arg Leu Phe Gly Lys Thr Ala Gly Arg Phe Asp Val Tyr Phe Ile Ser

100 105 110

Pro Gln Gly Leu Lys Phe Arg Ser Lys Ser Ser Leu Ala Asn Tyr Leu

115 120 125

His Lys Asn Gly Glu Thr Ser Leu Lys Pro Glu Asp Phe Asp Phe Thr

130 135 140

Val Leu Ser Lys Arg Gly Ile Lys Ser Arg Tyr Lys Asp Cys Ser Met

145 150 155 160

Ala Ala Leu Thr Ser His Leu Gln Asn Gln Ser Asn Asn Ser Asn Trp

165 170 175

Asn Leu Arg Thr Arg Ser Lys Cys Lys Lys Asp Val Phe Met Pro Pro

180 185 190

Ser Ser Ser Ser Glu Leu Gln Glu Ser Arg Gly Leu Ser Asn Phe Thr

195 200 205

Ser Thr His Leu Leu Leu Lys Glu Asp Glu Gly Val Asp Asp Val Asn

210 215 220

Phe Arg Lys Val Arg Lys Pro Lys Gly Lys Val Thr Ile Leu Lys Gly

225 230 235 240

Ile Pro Ile Lys Lys Thr Lys Lys Gly Cys Arg Lys Ser Cys Ser Gly

245 250 255

Phe Val Gln Ser Asp Ser Lys Arg Glu Ser Val Cys Asn Lys Ala Asp

260 265 270

Ala Glu Ser Glu Pro Val Ala Gln Lys Ser Gln Leu Asp Arg Thr Val

275 280 285

Ser Glu Glu Asn Ser Leu Val Lys Lys Lys Glu Arg Ser Leu Ser Ser

290 295 300

Gly Ser Asn Phe Cys Ser Glu Gln Lys Thr Ser Gly Ile Ile Asn Lys
305 310 315 320

Phe Cys Ser Ala Lys Asp Ser Glu His Asn Glu Lys Tyr Glu Asp Thr
325 330 335

Phe Leu Glu Ser Glu Glu Ile Gly Thr Lys Val Glu Val Val Glu Arg
340 345 350

Lys Glu His Leu His Thr Asp Ile Leu Lys Arg Gly Ser Glu Met Asp
355 360 365

Asn Asn Cys Ser Pro Thr Arg Lys Asp Phe Thr Gly Glu Lys Ile Phe
370 375 380

Gln Glu Asp Thr Ile Pro Arg Thr Gln Ile Glu Arg Arg Lys Thr Ser
385 390 395 400

Leu Tyr Phe Ser Ser Lys Tyr Asn Lys Glu Ala Leu Ser Pro Pro Arg
405 410 415

Arg Lys Ala Phe Lys Lys Trp Thr Pro Pro Arg Ser Pro Phe Asn Leu
420 425 430

Val Gln Glu Thr Leu Phe His Asp Pro Trp Lys Leu Leu Ile Ala Thr
435 440 445

Ile Phe Leu Asn Arg Thr Ser Gly Lys Met Ala Ile Pro Val Leu Trp
450 455 460

Lys Phe Leu Glu Lys Tyr Pro Ser Ala Glu Val Ala Arg Thr Ala Asp
465 470 475 480

Trp Arg Asp Val Ser Glu Leu Leu Lys Pro Leu Gly Leu Tyr Asp Leu
485 490 495

Arg Ala Lys Thr Ile Val Lys Phe Ser Asp Glu Tyr Leu Thr Lys Gln

500 505 510

Trp Lys Tyr Pro Ile Glu Leu His Gly Ile Gly Lys Tyr Gly Asn Asp

515 520 525

Ser Tyr Arg Ile Phe Cys Val Asn Glu Trp Lys Gln Val His Pro Glu

530 535 540

Asp His Lys Leu Asn Lys Tyr His Asp Trp Leu Trp Glu Asn His Glu

545 550 555 560

Lys Leu Ser Leu Ser

565

<210> 39

<211> 294

<212> PRT

<213> Deinococcus radiodurans

<400> 39

Gln Leu Gly Leu Val Cys Leu Thr Val Gly Pro Glu Val Arg Phe Arg
1 5 10 15

Thr Val Thr Leu Ser Arg Tyr Arg Ala Leu Ser Pro Ala Glu Arg Glu
20 25 30

Ala Lys Leu Leu Asp Leu Tyr Ser Ser Asn Ile Lys Thr Leu Arg Gly
35 40 45

Ala Ala Asp Tyr Cys Ala Ala His Asp Ile Arg Leu Tyr Arg Leu Ser
50 55 60

Ser Ser Leu Phe Pro Met Leu Asp Leu Ala Gly Asp Asp Thr Gly Ala
65 70 75 80

Ala Val Leu Thr His Leu Ala Pro Gln Leu Leu Glu Ala Gly His Ala
85 90 95

Phe Thr Asp Ala Gly Val Arg Leu Leu Met His Pro Glu Gln Phe Ile
100 105 110

Val Leu Asn Ser Asp Arg Pro Glu Val Arg Glu Ser Ser Val Arg Ala
115 120 125

Met Ser Ala His Ala Arg Val Met Asp Gly Leu Gly Leu Ala Arg Thr
130 135 140

Pro Trp Asn Leu Leu Leu Leu His Gly Gly Lys Gly Gly Arg Gly Ala
145 150 155 160

Glu Leu Ala Ala Leu Ile Pro Asp Leu Pro Asp Pro Val Arg Leu Arg

165 170 175

Leu Gly Leu Glu Asn Asp Glu Arg Ala Tyr Ser Pro Ala Glu Leu Leu

180 185 190

Pro Ile Cys Glu Ala Thr Gly Thr Pro Leu Val Phe Asp Ala His His

195 200 205

His Val Val His Asp Lys Leu Pro Asp Gln Glu Asp Pro Ser Val Arg

210 215 220

Glu Trp Val Leu Arg Ala Arg Ala Thr Trp Gln Pro Pro Glu Trp Gln

225 230 235 240

Val Val His Leu Ser Asn Gly Ile Glu Gly Pro Gln Asp Arg Arg His

245 250 255

Ser His Leu Ile Ala Asp Phe Pro Ser Ala Tyr Ala Asp Val Pro Gln

260 265 270

Ile Glu Val Glu Ala Lys Gly Lys Glu Glu Ala Ile Ala Ala Leu Arg

275 280 285

Leu Met Ala Pro Phe Lys

290

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12910

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :A01N 61/00, 37/18; C07H 21/02, 21/04
US CL :514/1, 2; 536/23.1, 23.2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 514/1, 2; 536/23.1, 23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
APS, STN, MEDLINE, BIOSIS CAPLUS, SCISEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	Database GenBank, GenBank Accession No. U78487, DAVEY et al., 'Direct Submission,' submitted 15 November 1996, Cancer Research Labs, Queen's University, Ontario, Canada, see sequence listing.	1-4 ------ 6
Y	DAVEY et al. The Fission Yeast UVDR DNA Repair Pathway is Inducible. Nucleic Acids Research. 1997, Vol. 25, No. 5, pages 1002-1008, especially page 1004.	6
X	Database GenBank, GenBank Accession No. D78571, TAKAO, M., 'Direct Submission,' submitted 29 November 1995, Tohoku University, Inst. of Development, Aging and Cancer, Miyagi 980, Japan, see sequence listing.	5

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 09 SEPTEMBER 1999	Date of mailing of the international search report 21 OCT 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer SHIN-LIN CHEN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/12910

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	TAKAO et al. Characterization of a UV Endonuclease Gene From the Fission Yeast *Schizosaccharomyces pombe* and Its Bacterial Homolog. Nucleic Acids Research. 1996, Vol. 24, No. 7, pages 1267-1271, especially page 1269.	7, 8, 10, 12-15, 17 ------ 9, 11, 16
Y	US 5,190,762 A (YAROSH) 02 March 1993, whole document, especially column 32.	14
Y	YONEMASU et al. Characterization of the Alternative Excision Repair Pathway of UV-damaged DNA in *Schizosaccharomyces pombe*. Nucleic Acids Research. 1997, Vol. 25, No. 8, pages 1553-1558, whole document.	16
Y, P	YOON et al. Processing of UV Damage In Vitro by FEN-1 Proteins as Part of an Alternative DNA Excision Repair Pathway. Biochemistry. 24 March 1999, Vol. 38, No. 15, pages 4809-4817, whole document.	16